VHF Air/Ground Communications for Air Traffic Control: A Decision Tree Approach to System Innovations

M 93B0000096V2
July 1993

Cheng-Hong Chen
James W. Howland
Robert I. Millar
Brian E. White
Warren J. Wilson

MITRE
Bedford, Massachusetts

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1993	

4. TITLE AND SUBTITLE
VHF Air/Ground Communications for Air Traffic Control: A Decision Tree Approach to System Innovations

5. FUNDING NUMBERS
DTFA01-93-C-00001

6. AUTHOR(S)
Cheng-Hong Chen, James W. Howland, Robert I. Millar, Brian E. White, Warren J. Wilson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The MITRE Corporation
202 Burlington Road
Bedford, MA 01730-1420

8. PERFORMING ORGANIZATION REPORT NUMBER
M 93B0000096V2

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
FAA

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Improvements to VHF air/ground communications for civil aviation in the 188-137 MHz aeronautical mobile frequency band are systematically explored by means of a decision tree approach. Seven individual papers analyze in detail the operational and technical factors involved in making these improvements. Basic tradeoffs between analog and digital modulation are discussed to frame the problem. Near-term improvements utilizing various analog modulations with closer channel spacing are first reviewed. Then far-term improvements employing a wide range of digital modulation and coding techniques are considered. Multiplexing methods - frequency, time, and code division - are discussed in detail. Methods for random access to a shared communications channel are compared, with emphasis upon real-time operation for air traffic control. Volume I provides an executive summary of this work, while Volume II presents the technical details.

14. SUBJECT TERMS
VHF air/ground communications, analog modulation, digital modulation, air traffic control

15. NUMBER OF PAGES
275

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

VHF Air/Ground Communications for Air Traffic Control: A Decision Tree Approach to System Innovations

M 93B0000096V2
July 1993

Cheng-Hong Chen
James W. Howland
Robert I. Millar
Brian E. White
Warren J. Wilson

Contract Sponsor FAA
Contract No. DTFA01-93-C-00001
Project No. 175DA
Dept. D050

Approved for public release;
distribution unlimited.

MITRE
Bedford, Massachusetts

Department Approval: Robert I. Millar

Robert I. Millar

MITRE Project Approval: Brian E. White

Brian E. White

ABSTRACT

Improvements to VHF air/ground communications for civil aviation in the 118-137 MHz aeronautical mobile frequency band are systematically explored by means of a decision tree approach. Seven individual papers analyze in detail the operational and technical factors involved in making these improvements. Basic tradeoffs between analog and digital modulation are discussed to frame the problem. Near-term improvements utilizing various analog modulations with closer channel spacing are first reviewed. Then far-term improvements employing a wide range of digital modulation and coding techniques are considered. Multiplexing methods — frequency, time, and code division — are discussed in detail. Methods for random access to a shared communications channel are compared, with emphasis upon real-time operation for air traffic control. Volume I provides an executive summary of this work, while Volume II presents the technical details.

ACKNOWLEDGEMENTS

The authors acknowledge the continued interactions with CAASD, particularly department F085, ATC Communications Systems. Special thanks are due Lisandro del Cid, Chris Moody, Dave Chadwick, and Amr ElSawy for their encouragement, project guidance, and technical suggestions. The involvement of our FAA sponsor, ASE-200, in this work has been very helpful, viz., Bruce Eckstein's program management oversight; Cindy Peak provided the original impetus for the decision tree approach. The interest of RTCA Special Committee 172 participants has also been stimulating.

Other members of the Bedford team deserve recognition for their contributions in preparing this document as well: Cho-Chau Li, Anne Doherty, Ginny Fitzgerald, Lynn Lacroix, and Nancy MacLeod.

Significant contributions to Section 6 of Volume II were made by Ivan La-Garde and Dave Snodgrass of Department D053.

TABLE OF CONTENTS FOR VOLUME II

SECTION		PAGE
4	Organization of the Report	4-1
5	Analog Versus Digital Modulation	5-1
5.1	Context	5-1
5.2	Background	5-1
5.3	Issues	5-3
5.4	Tradeoffs	5-5
5.5	Impact/Importance	5-21
5.6	Criteria for Decision	5-22
5.7	Connectivity/Relationship With Other Decisions	5-24
6	Analog Modulations	6-1
6.1	Context	6-1
6.2	Background	6-1
6.3	Issues	6-2
6.4	Tradeoffs	6-5
6.5	Impact/Importance	6-18
6.6	Criteria for Decision	6-19
6.7	Connectivity/Relationship with Other Decisions	6-21
7	Spectrum Utilization - Closer Channel Spacing	7-1
7.1	Context	7-1
7.2	Background	7-1
7.3	Issues	7-2
7.4	Tradeoffs	7-17
7.5	Impact/Importance	7-28
7.6	Transition	7-30
7.7	Criteria for Decision	7-32
7.8	Connectivity/Relationship with Other Decisions	7-33
8	Digital Modulations	8-1
8.1	Context	8-1
8.2	Background	8-1
8.3	Issues	8-3
8.4	Tradeoffs	8-13
8.5	Impact/Importance	8-31
8.6	Transition	8-32
8.7	Criteria for Decision	8-33
8.8	Connectivity/Relationship with Other Decisions	8-36

TABLE OF CONTENTS FOR VOLUME II (CONTINUED)

SECTION			PAGE
9 Coding Techniques			9-1
9.1	Context		9-1
9.2	Background		9-1
9.3	Issues		9-3
9.4	Tradeoffs		9-11
9.5	Impact/Importance		9-29
9.6	Transition		9-29
9.7	Criteria for Decision		9-30
9.8	Connectivity/Relationship with Other Decisions		9-32
10 Multiplexing			10-1
10.1	Context		10-1
10.2	Background		10-1
10.3	Issues		10-2
10.4	Tradeoffs		10-14
10.5	Impact/Importance		10-21
10.6	Criteria for Decision		10-25
10.7	Connectivity/Relationship with Other Decisions		10-26
11 Random Access			11-1
11.1	Context		11-1
11.2	Background		11-1
11.3	Issues		11-11
	11.3.1	Pure and Slotted ALOHA	11-12
	11.3.2	Reservation ALOHA	11-16
	11.3.3	Collision Resolution Algorithms	11-18
	11.3.4	Mini-Slotted Alternating Priorities	11-27
	11.3.5	Non-Persistent and p-Persistent CSMA	11-32
	11.3.6	Virtual Time CSMA	11-35
11.4	Tradeoffs		11-41
	11.4.1	Pure ALOHA	11-44
	11.4.2	Slotted ALOHA	11-44
	11.4.3	Reservation ALOHA	11-44
	11.4.4	Collision Resolution Algorithms	11-45
	11.4.5	Mini-Slotted Alternating Priorities	11-45
	11.4.6	p-Persistent CSMA	11-45
	11.4.7	Virtual Time CSMA	11-46
	11.4.8	Summary	11-47

SECTION			PAGE
11.5	Impact/Importance		11-47
	11.5.1	Best RA Scheme Performers	11-47
	11.5.2	Assessment of Lower Performance RA Schemes	11-54
11.6	Transition		11-55
11.7	Criteria for Decision		11-57
11.8	Connectivity/Relationship with Other Decisions		11-60
List of References			RE-1
Appendix A			A-1
Appendix B			B-1
Glossary			GL-1
Distribution List			DI-1

LIST OF FIGURES

FIGURE		PAGE
6-1	Amplitude Modulation (Double Sideband Transmitted Carrier)	6-7
6-2	Double Sideband Suppressed Carrier	6-8
6-3	Quadrature Amplitude Modulation	6-8
6-4	Single Sideband	6-9
6-5	Vestigial Sideband	6-9
6-6	Frequency Modulation	6-10
6-7	Narrowband Frequency Modulation	6-10
7-1	Insertion Loss (S21)	7-13
7-2	Return Loss (S11)	7-14
7-3	Group Delay	7-15
7-4	Phase Response	7-16
8-1	Channel Signal Constellations and Gray Codes	8-5
8-2	Amplitude Mask Requirement for A-QPSK Signals	8-9
8-3	Phase Mask Requirement for A-QPSK Signals	8-10
8-4	Baseband Equivalent Power Spectra (Baseband Bandwidth Used)	8-19
8-5	Fractional Out-of-Band Power for Various Modulation Schemes	8-21
8-6	Channel Bandwidth-Power Tradeoff for Various Modulation Schemes Without Coding	8-34
9-1a	Convolutional Encoder with Constraint Length K and Rate k/n	9-9
9-1b	A Rate 1/2, Constraint Length 7 Convolutional Encoder	9-9
9-2	P_{be} Versus E_b/N_0 over a Gaussian Channel for Several Half-Rate Binary and High-Rate RS Block Codes	9-13
9-3	P_{be} Versus E_b/N_0 over a Gaussian Channel for Several High-Rate RS Block Codes	9-14
9-4	Packet (Undetected) Error Rate Versus Channel Bit-Error-Rate	9-21
9-5	Throughput Efficiency of Three Basic ARQ Schemes	9-25
10-1	Air Traffic Control Radio Spectrum	10-7
11-1a	"Acceptable" Region for Packet Length (L) and Information Rate (R); Altitude = 72,000 ft	11-6

FIGURE		PAGE

11-1b	"Acceptable" Region for Packet Length (L) and Information Rate (R); Altitude = 18,000 ft	11-7
11-1c	"Acceptable" Region for Packet Length (L) and Information Rate (R); Altitude = 4,500 ft	11-8
11-2	Theoretical Throughput vs. Offered Channel Traffic	11-14
11-3a	Delay vs. Channel Loading for Pure ALOHA	11-15
11-3b	Delay vs. Channel Loading for Slotted ALOHA	11-15
11-4	Slotted ALOHA, TDMA, and First-In, First-Out (FIFO) Reservation: Delay Throughput Tradeoff for 50 Users and Single-Packet Messages in a Satellite Environment [11]	11-17
11-5	Collision Resolution Example Using the Basic Algorithm	11-19
11-6	Bounds on Serial-Tree Delay [12]	11-22
11-7a	Throughput vs. Collision Sensing Response (S < Inverse of Coefficient of N in Equation (13))($\varepsilon = 0.01$)	11-26
11-7b	Throughput vs. Collision Sensing Response (S < Inverse of Coefficient of N in Equation (13))($\varepsilon = 0.1$)	11-26
11-8	Average Packet Delay vs. Traffic Loading	11-31
11-9	CSMA and ALOHA: Throughput-Delay Tradeoffs from Simulation (Normalized Propagation Delay a = 0.01)	11-34
11-10	Throughput vs. Traffic Loading for Virtual Time CMSA	11-39
11-11	Delay for Synchronous Virtual Time CSMA	11-42
11-12a	Approximate Regions of Support to Real-Time Operations by the Most Capable Random Access Schemes	11-49
11-12b	Approximate Regions of Support to Real-Time Operations by the Most Capable Random Access Schemes	11-50
11-13	Qualitative Asessment of Random Access Algorithms with Respect to Their Feasibility of Supporting Real-Time Operations	11-59

LIST OF TABLES

TABLE		PAGE
5-1	Analog Versus Digital Tradeoffs I	5-6
5-2	Analog Versus Digital Tradeoffs II	5-7
5-3	Analog Versus Digital Tradeoffs III	5-8
5-4	Analog Versus Digital Tradeoffs IV	5-9
6-1	Comparison of Several Analog Modulations	6-6
8-1	Quantitative Comparison of Several Advanced Digital Modulations (Additive White Gaussian Noise)	8-15
8-2	Various Bandwidths for Some Digital Modulation Schemes (Normalized to Data Rate, 1/T)	8-23
8-3	Quantitative Comparison of Several Advanced Digital Modulations (Rayleigh Fading Channel)	8-29
9-1	Bandwidth Efficency and Power Efficiency of Block Codes with OQPSK Modulation	9-18
9-2	Bandwidth Efficiency and Coding Gain for Convolutional Codes wtih OQPSK Modulation and Viterbi Decoding	9-19
9-3	Bit-Error-Rate Performance of Rate 1/2, K=7, Hard Decision Convolutional Code with Viterbi Decoding in Rayleigh Fading	9-27
10-1	Comparison of Multiplexing Techniques	10-22
11-1	Channel Utilization, U, for Several Values of Traffic Loading, G ($a=0.1$; $\delta/T=0.05$, i.e., $v=0.1$)	11-32
11-2	Extremes of Maximum Propagation Delay, τ, for Three Altitude Regimes (h is altitude, in feet; c is speed of light in feet per second)	11-43
11-3	Packet Durations, T=L/R, of Interest and the Real-Time Thresholds	11-43
11-4	Characteristics of the RA Algorithms Discussed With Emphasis on Their Ability to Support Real-Time Operation	11-48

SECTION 4

ORGANIZATION OF THE REPORT

This report is divided into two volumes. Volume I presents the motivation for the work, describes a decision tree approach to system innovations, and explains which branches of the decision tree were selected for detailed investigation. Then an executive summary of each of seven individual decision tree papers is given. Finally, some conclusions are drawn about the utility of the decision tree approach and recommendations for further work are made.

Volume II is a compilation of seven individual decision tree papers. Each paper provides some background about where it fits in the larger picture, delineates the relevant technical and operational issues, discusses engineering tradeoffs in some detail, summarizes the impact and importance of these tradeoffs, reviews the criteria for decision, and provides the connectivity and relationship with other branches of the decision tree. The seven decision tree papers and their authors are:

- Analog versus Digital Modulation (Millar)
- Analog Modulations (Millar)
- Spectrum Utilization — Closer Channel Spacing (Howland)
- Digital Modulations (Chen)
- Coding Techniques (Chen)
- Multiplexing (Wilson)
- Random Access (White)

SECTION 5

ANALOG VERSUS DIGITAL MODULATION

5.1 CONTEXT

A systematic process for examining technical alternatives for improved air/ground communications in air traffic management has been established. A decision tree structure is shown in Appendix A that attempts to organize various alternatives in a top-down hierarchy. This provides a framework for evaluating potential solutions that can be represented by paths through the decision tree.

This paper addresses a particular alternative in the decision tree, namely, branchpoint 4.1, Advanced Modulation/Coding. The organization of this paper is as follows: Background, Issues, Tradeoffs, Impact/Importance, Criteria for Decision, and Connectivity/Relationship with Other Decisions.

5.2 BACKGROUND

The decision tree considers a wide range of possible improvements to VHF air/ground communications. These fall into two general categories:

1) Improving the current air/ground communications system while maintaining the same concept of operations. This includes increasing the number of voice channels available in the 118-137 MHz aeronautical mobile frequency band, reducing susceptibility to interference and fading, improving speech intelligibility, increasing link availability, etc.

2) Adding new capabilities to the current air/ground communications system that change the concept of operations. These include automated frequency changes and computer-assisted handoff of aircraft from one air traffic controller to another, different channel access methods, provisions for emergency break-in to a busy channel, off-loading routine voice communications onto data communications channels, etc.

Choosing between analog and digital modulation to improve VHF air/ground communications can be viewed in the context of these two categories of system improvements. Many of the improvements in the first category can be obtained with either analog or digital modulation, but many of the improvements in the second category are more readily provided with digital modulation. Section 6 discusses the capabilities of seven specific types of analog modulation, particularly for increasing the number of voice channels in the 118-137 MHz band. Section 8 discusses the capabilities of many types of digital modulation, addressing particularly the spectral efficiency of different digital communications signaling waveforms; i.e., how many bits/sec of digital data can be transmitted in a given radio-frequency bandwidth in Hertz. Section 10 relates the spectral efficiencies of these digital modulations to various signal multiplexing methods, in order to determine how many digital voice channels can be provided in the 118-137 MHz frequency band.

This paper will look at how the basic choice of analog or digital modulation, branchpoint 4.1 on the decision tree, relates to many other branchpoints, including 4.2, channel access; 4.3.2, multiplexing; 2.1.1.1, air/ground link redundancy; 2.1.3, busy channel handling; 2.3, workload mitigation; and 2.4, data link applications; among others. The choice of analog versus digital modulation will be viewed in the context of the overall air/ground communications systems design, relying upon the following sections to provide the detailed technical tradeoffs among individual modulation waveforms of each type, analog and digital.

Tradeoffs between analog and digital modulation in this paper consider only waveforms which fit into a channel bandwidth of 25 kHz or less. Spread spectrum waveforms are discussed in section 10 in terms of code division multiplexing.

Finally, the question of analog versus digital modulation is being discussed in other forums, where various proposals are being considered for VHF air/ground communications improvements. At International Civil Aviation Organization (ICAO) meetings, European nations have proposed reducing the frequency channelization of the 118-137 MHz band from 25 kHz to 12.5 kHz as a means of providing more channels with the currently used amplitude modulation (AM). RTCA Special Committee 172 is considering possible improvements to VHF air/ground communications in the 118-137 MHz band. Working Group 1 is considering long-term operational requirements and new system architectures, including the

use of different modulation techniques. Working Group 2 is preparing Minimum Aviation System Performance Standards to define the signal-in-space characteristics for VHF digital data communications, including compatibility with digital voice techniques.

5.3 ISSUES

Improved VHF air/ground communications capability can be defined by specific technical and operational criteria. The first five items below represent improved capabilities that are consistent with the current concept of operations. The next six items represent improved capabilities that change the concept of operations to a greater or lesser degree. The last four items are implementation issues relating to the practicality of new modulations. The 15 criteria chosen are:

1. Number of voice channels per 25 kHz.

2. Power efficiency, defined as the relative transmitter power required per voice channel.

3. Susceptibility to interference, including both cochannel and adjacent-channel interference.

4. Performance degradation caused by both slow and fast fading in the VHF air/ground channel.

5. Speech intelligibility, including both normal (single-talker) and overload (two-talker) conditions on a voice channel.

6. Data transmission capability through a voice radio.

7. Compatibility of modulation with multiplexing methods, including frequency, time, and code division.

8. Compatibility of modulation with channel access methods, including fixed assigned, demand assigned, and random access.

9. Evolutionary growth capability, particularly, the ability to provide more voice channels when lower data rate vocoders become available.

10. Compatibility of modulation with channel availability and message integrity improvements, including frequency and path diversity, busy channel override, and automatic repeat request (ARQ) of messages.

11. Compatibility of modulation with advanced user features, such as automated handoff, group addressing, and air-to-air selective routing.

12. Need for Doppler tracking and coherent demodulation.

13. Radio design complexity required to transmit and receive the modulation.

14. Technical maturity and design risk of the modulation.

15. Backward compatibility with current AM radios.

To compare analog and digital modulation against these 15 criteria, we have chosen two representative modulations of each type. A detailed evaluation of seven analog modulations in section 6 concludes that amplitude modulation (AM) and single sideband (SSB) are appropriate choices for improving VHF air/ground communications. AM is a simple modulation that can be demodulated with an envelope detector, while SSB is a more complex modulation requiring a coherent detector for demodulation, but SSB can provide twice as many voice channels as AM in a given radio-frequency bandwidth.

Because there are so many types of digital modulation (see section 8), there are no two that appear clearly best at this point. Therefore, we have chosen as representative the two modulations currently under consideration by Working Group 2 of RTCA Special Committee 172 as a signal-in-space for VHF digital data communications, including compatibility with digital voice techniques. These two digital modulations are 4-ary offset quadrature amplitude modulation (4-OQAM) and 16-ary offset quadrature amplitude modulation (16-OQAM). The 4-OQAM digital modulation has a nearly constant envelope and can be demodulated by

a phase-comparison demodulator, while 16-OQAM has varying amplitude levels and requires a coherent detector for demodulation, but 16-OQAM has twice the data rate of 4-OQAM for a given radio-frequency bandwidth. Thus, for both the analog and digital modulations, we have chosen a simple limited performance waveform and a more complex higher performance waveform to illustrate the range of capabilities available with each type of modulation.

5.4 TRADEOFFS

Tables 5-1 to 5-4 summarize the performance of analog and digital modulation against the 15 operational and technical criteria described above. Tradeoffs between analog and digital modulation will be discussed in turn for each criterion.

<u>Number of Voice Channels</u>

Currently, a 25 kHz channel spacing is specified for the 118-137 MHz aeronautical mobile frequency band in the United States and Europe, providing 760 channels. Some parts of the world with less air traffic employ *50 kHz* or *100 kHz channel spacing*, providing 380 or 190 channels, respectively. Since the actual occupied bandwidth of an AM voice signal is only 7 kHz, channel spacing of AM radios could be reduced to 12.5 kHz or even to 10 kHz.

Table 5-1. Analog vs. Digital Tradeoffs I		
Criterion	**Analog Modulation**	**Digital Modulation**
Number of voice channels per 25 kHz	2 or 2.5 channels with AM. 4 or 5 channels with SSB. Cosite interference may limit number of useful channels.	2 or 3 channels with 4-OQAM and 7200-9600 b/s vocoders. 3 to 5 channels with 16-OQAM and 4800-6500 b/s vocoders.
Power efficiency	AM is poor. SSB is 5 to 10 dB better.	4-OQAM is very good. 16-OQAM is good.
Susceptibility to interference	C/I at least 14 dB for AM. SSB may be better, but requires analysis.	C/I about 10-15 dB for digital modulation.
Performance in fading	AM and SSB are both affected.	Some resistance to fast fading with equalization and error correction coding.

Table 5-2. Analog vs. Digital Tradeoffs II		
Criterion	**Analog Modulation**	**Digital Modulation**
Speech intelligibility	AM and SSB are good at a high signal-to-noise ratio. Quasi-linear conferencing exists with SSB.	Intelligibility is determined by vocoder — better at higher bit rates. Cockpit acoustic noise is a concern.
Data transmission capability	Voiceband signaling provides a limited data rate.	Data transmission at 20,000-40,000 b/s is possible.
Compatibility with multiplexing	Frequency division multiplexing (FDM) only.	Compatible with all multiplexing methods.
Compatibility with channel access	Compatible with FDMA, polling, and trunking.	Compatible with all channel access methods.

Table 5-3.
Analog vs. Digital Tradeoffs III

Criterion	Analog Modulation	Digital Modulation
Evolutionary growth capability	AM and SSB are the end products of a mature art.	Lower bit rate vocoders can provide more channels. Growth provisions can be built into TDMA.
Compatibility with channel availability and message integrity improvements	Path diversity and busy channel overrides can use in-band signaling. Frequency diversity and ARQ are difficult to provide.	Flexible implementation of frequency and path diversity, busy channel overrides, and ARQ is possible.
Compatibility with advanced user features	Features can be implemented by in-band signaling, with some interruption of voice.	Automated handoff and group addressing can be done with TDMA signaling channels, with no interruption of voice.
Doppler tracking and coherent demodulation	Not needed with AM. SSB needs frequency tracking to 20 Hz accuracy.	4-OQAM can use differential-phase demodulation. 16-OQAM requires coherent demodulation.

Table 5-4. Analog vs. Digital Tradeoffs IV		
Criterion	**Analog Modulation**	**Digital Modulation**
Radio design complexity	AM has simple circuitry. SSB requires linear amplifiers plus pilot tone tracking loop.	Vocoder complexity is transparent to user. Digital radios are more complex than analog radios, particularly with coherent demodulation.
Technical maturity and development risk	AM and SSB are mature technologies with low risk.	4-OQAM, GMSK, and $\pi/4$-QPSK have low-to-moderate risk. 16-OQAM is less proven for the VHF air-ground channel.
Backward compatibility with current AM radios	AM with 12.5 kHz spacing has the easiest backward compatibility. SSB radio would need two different modes.	Digital radio can also have an AM mode. Additional circuitry is needed: modulators, demodulators, IF filters, and AGC.

This would require aircraft and ground stations to be equipped with new AM radios with more accurate local oscillators and better receiver selectivity (see section 7). Single sideband uses half the radio frequency bandwidth of AM, which is sometimes called double sideband transmitted carrier. Therefore, by converting to SSB radios, a 6.25 kHz or even 5 kHz channel spacing could be employed by VHF air-ground communications. Thus, 2 or 2.5 voice channels per 25 kHz are possible with AM and 4 or 5 voice channels are possible with

SSB. As a caveat, if the additional closely-spaced channels are used at locations where cosite interference is a limiting factor, the useful number of channels will be less than the nominal number.

Digital modulation can also provide several voice channels per 25 kHz (see section 10). Using 4-OQAM (or comparable waveforms like π/4-QPSK and Aviation-QPSK) at baud rates that ensure low adjacent channel interference and employing 7200-9600 b/s vocoders, two or three voice channels per 25 kHz can be provided by time-division multiplexing (TDM). More channels can be provided by increasing the bit rate in a 25 kHz channel. This requires either (1) increasing the baud rate for 4-OQAM and relaxing the adjacent-channel interference specification, or (2) using 16-OQAM, which doubles the bit rate as compared to 4-OQAM at the same baud rate. With a higher bit rate and employing 4800-6500 b/s vocoders, three to five TDM voice channels become available in a 25 kHz bandwidth. One sees that either analog or digital modulation can provide significantly more voice channels for VHF air/ground communications.

Power Efficiency

Power efficiency of AM is poor because most of the transmitted power is used to transmit the radio-frequency carrier and only a small part is used to transmit the modulation sidebands that convey the speech information. Audio signal-to-noise ratio is lower than carrier-to-noise ratio by 5 to 10 dB, typically. The exact amount depends upon the modulation index, which constantly varies for speech. Single sideband offers better power efficiency than AM because audio signal-to-noise ratio is equal to signal-to-noise ratio at intermediate frequency (IF) on an average power basis. No transmitter power is wasted in transmitting a carrier (but a small amount is needed to transmit a pilot tone for Doppler tracking on the air-ground channel). With all analog modulations, audio signal-to-noise ratio increases monotonically with increasing transmitter power.

The power efficiency of quaternary digital modulations like 4-OQAM is very good. When the energy per bit to noise power spectral density ratio, E_b/N_0, exceeds 8 dB, digital voice becomes essentially error free. When a high audio signal-to-noise ratio is required on a voice channel, digital modulation can usually provide it at a lower transmitter power than an analog modulation because of this threshold effect. A 16-ary modulation like 16-OQAM

gives up some power efficiency to double its spectral efficiency; e.g., approximately a 4 dB increase in average transmitter power per voice channel is needed for 16-OQAM to obtain the same bit error rate as 4-OQAM.

Susceptibility to Interference

Cochannel interference has somewhat the same effect as noise upon analog modulation. A high signal-to-interference ratio is required for high quality speech, but intelligible speech can be obtained with AM at around 14 dB signal-to-interference ratio. It is possible that SSB would be intelligible at a lower signal-to-interference ratio, but this requires analysis of the effects of interference upon Doppler tracking of a pilot tone.

Digital modulation tends to be resistant to cochannel interference. Once signal-to-interference ratio gets above 10-15 dB, digital voice becomes essentially interference free. Gaussian minimum shift keying (GMSK) is particularly resistant to cochannel interference. Digital waveforms using multiple amplitude levels like 16-OQAM are more susceptible to cochannel interference, but can still provide higher quality speech at a given signal-to-interference ratio than analog waveforms.

Adjacent channel interference is currently a manageable problem with AM because the occupied bandwidth of 7 kHz is small compared to the 25 kHz channel width. Were the channel width to be reduced to 12.5 kHz or 10 kHz, better local oscillators and narrower IF bandwidths would be required to keep adjacent channel interference low (see section 7). Similar design considerations would apply to SSB radios operating with channel widths of 6.25 kHz or 5 kHz.

With digital modulations, adjacent channel interference is caused by modulation sidebands; i.e., spillover of the frequency spectrum of the digital waveform into adjacent channels. Quantitative data for different digital modulations is summarized in section 8. In general, the need to keep adjacent channel interference 40 to 70 dB below the signal power in the assigned channel limits the usable data rates of digital modulation in a 25 kHz channel to 20,000 - 30,000 b/s with 4-OQAM and about 40,000 b/s with 16-OQAM.

Cosite interference among multiple transmitters and receivers at a ground station or on a large aircraft is another factor limiting VHF air/ground communications performance. Intermodulation product interference increases rapidly with the number of simultaneous transmissions; i.e., the number of cosite carriers present. For example, the number of two-frequency third-order intermodulation products generated by n transmitters is n(n-1). These are generally the strongest of the many intermodulation products generated by nonlinearities in transmitters, receivers, and other components. Frequency management at a ground station usually involves selecting carrier frequencies so that one does not try to receive on a frequency which is a two-frequency third-order intermodulation product of any pair of transmitters. Following this rule becomes difficult if there are many transmitters and receivers at a ground station. In addition, other causes of cosite interference, such as crossmodulation and receiver desensitization get worse as the number of simultaneous transmissions increase. The result is that the useful number of channels at some ground stations increases more slowly than the increase in the nominal number of channels when narrower channel widths are used. Narrower channel widths are the means of providing more voice channels with analog modulation, but not with digital modulation, which provides more voice channels by time division multiplexing. Cosite interference, therefore, is more of a limiting factor with analog modulation than with digital modulation.

Performance in Fading

VHF air/ground radio links are subject to both slow and fast fading. Slow fading is caused by such physical phenomena as specular earth reflection, foliage attenuation, atmospheric refraction, atmospheric multipath, and airborne antenna shadowing. Slow fading can be deep, with a signal loss of 10 to 20 dB or more, with the fades lasting for seconds or even minutes. All analog and digital modulations are affected by slow fading that causes a loss in signal-to-noise ratio at the radio receiver; this loss can be catastrophic on weak signals. With strong signals, slow fading is compensated for by receiver automatic gain control (AGC) action.

Fast fading is caused by such physical phenomena as diffuse earth reflection, particularly from buildings, airframe reflections and diffraction, and rotor-wing aircraft blade effects. Fast fading is not as deep as slow fading, with signal loss usually less than 10 dB. The bandwidth, or Doppler spread, of fast fading can be large, however, up to tens of Hertz for

air/ground links in the 118-137 MHz band. If the bandwidth of fast fading extends to audible frequencies, the resulting fluctuations in signal amplitude may cause distortion of the audio for such analog modulations as AM and SSB, even when the received signal-to-noise ratio is high. If one makes the receiver AGC fast enough to respond to fast fading, then the AGC will cancel out lower frequency components in the audio. There is a tradeoff in the receiver AGC design between mitigating fast fading effects and preserving lower audio frequencies.

Digital modulations which are variants of quadrature phase-shift keying, including 4-OQAM, $\pi/4$-QPSK, and Aviation QPSK, have some resistance to fast fading because the four points in the signal constellation have the same amplitude and well-separated phase angles. Thus, signal amplitude fluctuations produce no errors and phase fluctuations must be large to produce errors. With 16-OQAM, the 16 points in the signal constellation are closer together, so that amplitude fluctuations can produce errors, as can smaller phase fluctuations. The 16-OQAM modulation may still have better performance in fading than analog modulation, but this requires more analysis before drawing definite conclusions. Resistance to fast fading can be improved with channel equalization, error correction coding (see section 9), and bit interleaving, but these measures are not particularly useful against slow fading because the fade duration often exceeds the effective time-span of the equalization, coding, and interleaving.

Speech Intelligibility

With a single talker on a voice channel, the normal operating condition, analog modulations such as AM and SSB can provide noisy but intelligible speech at an IF signal-to-noise ratio around 10-15 dB, but high quality speech requires a much higher signal-to-noise ratio. Digital modulations have a threshold: when E_b/N_0 is above 8 dB for 4-OQAM and above 12 dB for 16-OQAM, no noticeable degradation in speech intelligibility is produced by the radio link. Intelligibility may be limited by the characteristics of the vocoders, however. In particular, many vocoders in the 4800-6500 b/s range have not been tested in the presence of cockpit acoustic noise.

In some cases, two talkers will inadvertently transmit on the same channel at the same time, a phenomenon termed "walk-on." With AM, carrier heterodyning produces an audible tone with substantial audio distortion. A strong point of SSB is its quasi-linear response to

two talkers on the same channel, resulting in graceful degradation, rather than catastrophic failure, in this situation. With slow aircraft, the Doppler shifts of the two talkers will produce small pitch shifts in their audio, but some degree of intelligibility will remain. The pilot tone tracking loop that tracks signal Doppler shift does not have to work perfectly with SSB to produce intelligible audio. With fast aircraft, however, the audio pitch shifts of the two talkers will be larger and intelligibility poorer, because the pilot tone tracking loop can track only one of the two Doppler-shifted pilot tones.

With digital modulation, two talkers on the same channel will interfere with each other at each receiver. If the two signals are of comparable amplitude, neither will be intelligible. If one is much stronger than the other, the stronger signal will capture the digital demodulator and be understood, with the weaker signal acting as interference. In a well-designed digital communication system, it is possible to provide "collision detection" capabilities that detect two talkers on a channel and send a message over a service channel to shut down one of the offending transmitters. This eliminates the "walk-on" problem.

Data Transmission

Digital data can be transmitted and received by analog voice radios with voiceband signaling. This technique, used by telephone modems, requires digital data to be represented by waveforms whose bandwidth is less than or equal to that of speech. The bandwidth restriction limits the data rates of voiceband signaling to a few thousand bits per second, much less than the data rates that can be sent over a digital radio.

A digital radio can be used to transmit and receive either digital voice or digital data. Real-time digital voice requires a limited channel delay, but can tolerate bit error rates as high as 1%. Digital data can tolerate longer channel delays, but requires bit error rates in the range of 10^{-5} to 10^{-9}. These can be achieved by combining digital modulation with both error correcting codes and error detecting codes, as discussed in section 9. Channel data rates from 20,000 to 40,000 b/s are possible in a 25 kHz channel.

Compatibility with Multiplexing

Analog modulations such as AM and SSB are compatible with frequency division multiplexing (FDM). Radio communications in the HF, VHF, and UHF bands have employed FDM for over 60 years, and the analog radio art has developed around FDM design principles. Analog modulation is not compatible with time division multiplexing (TDM) or with most forms of code-division multiplexing (CDM). Analog modulation can be combined with slow frequency-hopping, but this form of CDM is not well suited to VHF air/ground communications, as shown in section 10.

Digital modulation is compatible with FDM, TDM. and CDM. Use of digital voice for air-ground communications allows a system designer to use any of these multiplexing methods. This design flexibility is a strong advantage of digital modulation.

Compatibility with Channel Access Methods

Analog modulation is compatible with only a few of the channel access methods shown under branchpoint 4.2 of the decision tree. Of the fixed assigned channel access methods (branchpoint 4.2.1) only frequency division multiple access (FDMA) and slow frequency hopping code division multiple access (CDMA) can be combined with analog modulation in a system design. Of the demand assigned channel access methods (branchpoint 4.2.2), FDMA, slow frequency hopping CDMA, polling, and trunked radio can be combined with analog modulation. Of the random access methods (branchpoint 4.2.3), only FDMA and slow frequency hopped CDMA can be combined with analog modulation. Section 11 describes various random access methods in detail. Analog modulation is not compatible with time division multiple access (TDMA), most forms of CDMA, token ring methods, carrier sense multiple access (CSMA), the ALOHA random access algorithms, or with various conflict resolution algorithm (CRA) procedures.

Digital modulation can be combined with all of the channel access methods shown under branchpoint 4.2 of the decision tree. Many of these channel access methods, such as TDMA, token ring, CSMA, and ALOHA, were specifically designed to be used with digital modulation.

Evolutionary Growth

Analog modulations such as AM and SSB are the end products of a mature art. As mentioned earlier, AM radios can provide 2 or 2.5 channels per 25 kHz by channel splitting and SSB radios can provide 4 or 5 channels. Further increases in the number of channels would be difficult with analog modulation, because the resulting AM channel bandwidths are only a little larger than twice the bandwidth of speech and the resulting SSB channel bandwidths are only a little larger than the speech bandwidth.

Research into speech encoding devices is producing better vocoders every year. Recent developments in improved multi-band excitation (IMBE) and code excited linear prediction (CELP) vocoders operating at 4800-6500 b/s are very promising and even lower bit rate vocoders are on the horizon. If these anticipated advances occur, the lower bit rates can be exploited to provide more than 4 or 5 voice channels per 25 kHz. Thus, VHF air/ground radios employing digital modulation have an inherent growth capability. With advance planning, digital radios designed for vocoders in the 4800-9600 b/s range could easily be changed over to use lower bit rate vocoders.

Compatibility with Channel Availability Improvements

Channel availability can be improved by employing frequency diversity or path diversity on air/ground radio links (branchpoint 2.1.1.1), by providing busy channel overrides, and by employing an automatic repeat request (ARQ) feature. Although frequency diversity can be provided with analog modulation in an FDMA system design, it may lead to inefficient use of frequency channels. Path diversity and busy channel overrides can be provided in an analog radio by using in-band or subcarrier signaling for system switching and control. The difficulty in providing ARQ with analog modulation is the fact that there is no clear-cut criterion for repeating a message. With digital modulation and error detection coding, a failure to decode a message provides such a criterion. Digital modulation also provides great flexibility in implementing frequency diversity, path diversity, and busy channel overrides. Some channel access methods designed for digital modulation include ARQ in their protocols.

Compatibility with Advanced User Features

Examples of advanced user features are automated frequency changes, computer-assisted handoff of aircraft from one air traffic controller to another, allowing a controller to address a group of aircraft, and selective routing of air-to-air messages. With analog voice radios, such features can be implemented with signaling and control channels on separate frequencies in an FDMA design or could use in-band signaling over the voice channels. In the latter case, some interruption of the voice signal would occur.

In a digital radio, advanced user features would also require signaling and control channels, which are readily provided in a TDMA design by assigning a fraction of the time slots per frame to signaling and control. A TDMA communications system design is particularly efficient at allocating just the required data rate to each function without wasting channel capacity on low data rate functions like signaling and control, as might occur in an FDMA design with fixed channel bandwidths. Thus, although advanced user features can be provided in analog radios, digital radios provide greater flexibility in allocating channel capacity between signaling and control and digital voice. Digital radios also do not require any interruption of voice for signaling and control.

Doppler Tracking and Coherent Demodulation

No doppler tracking is required with AM radios and a simple envelope detector suffices for voice demodulation. One makes the radio IF bandwidth a little wider than the double-sideband signal bandwidth of 7 kHz, to accommodate both Doppler shift of the received signal and local oscillator drift that together should not exceed 1 kHz at VHF. AM signal detection with an envelope detector is insensitive to this frequency shift.

A single sideband radio requires Doppler tracking and coherent demodulation. A carrier frequency must be generated in the receiver, to an accuracy of about 20 Hz for voice communications. Since this accuracy is much less than Doppler shift and local oscillator drift at VHF, it is necessary to transmit a pilot tone with the SSB signal and to frequency track it in the SSB receiver. Precise phase tracking of the pilot tone is not required for voice communications.

Quaternary-phase digital modulations, such as 4-OQAM, $\pi/4$-QPSK, and Aviation QPSK, can be demodulated coherently. This requires carrier phase tracking of both Doppler shift and local oscillator offsets between transmitter and receiver to provide an in-phase reference signal against which to compare the phase of the received signal. Coherent demodulation maximizes power efficiency. Alternatively, quaternary-phase digital modulations can be demodulated by phase-comparison demodulators, which make symbol decisions by comparing the phase of each received symbol with the phase of the previous symbol. Phase comparison demodulators require no Doppler tracking, but give up to 2 to 3 dB in power efficiency compared to coherent demodulators.

Digital modulations like 16-OQAM, with signal constellation points more closely spaced, require Doppler tracking and coherent demodulation to achieve desired bit error rates at reasonably low signal-to-noise ratios. With coherent demodulation, 16-OQAM requires only 4 dB more transmitter power per voice channel than 4-OQAM, in return for doubling the data rate in a 25 kHz channel.

Radio Design Complexity

An AM radio is simple to design and manufacture. It does require a linear power amplifier to transmit the peaks of speech signals and a linear receiver with AGC to receive the AM signal. An SSB transmitter requires a linear power amplifier. Linearity requirements for SSB are more stringent than those for AM because of the higher peak-to-average ratio of an SSB signal. Automatic level control (ALC) design for the power amplifier is more difficult than for AM; unlike the AM carrier, the SSB pilot tone may not have sufficient amplitude to drive an ALC. An SSB receiver needs a pilot tone tracking loop to remove the local oscillator offset between transmitter and receiver as well as the doppler shift and, furthermore, the performance of this frequency tracking loop is critical to audio quality and signal intelligibility. Tradeoffs exist between acquisition time of the tracking loop and frequency tracking accuracy. The receiver RF, IF, and audio amplification chain must be linear to accommodate signal peaks. Although an SSB transceiver is more complex than an AM transceiver, industry has experience building SSB radios in production quantities.

Digital radios employing phase-comparison detectors for quaternary-phase digital modulations are only a little more complex than analog radios. Transmitter power amplifiers must be linear since waveforms like 4-OQAM, $\pi/4$-QPSK, and Aviation QPSK do not have a constant envelope after premodulation filtering to reduce adjacent-channel interference. Limiting in the power amplifier will tend to restore a constant envelope, spreading the skirts of the modulation spectrum and resulting in an increase in spectrum energy in adjacent channels. Digital radios employing 16-OQAM or similar waveforms are more complex, since they must have coherent demodulators, linear power amplifiers, linear receivers, accurate carrier phase tracking, and digital signal processors to equalize channel fading. All digital radios need a vocoder. Vocoders in the 4800-6500 b/s range employ some variant of linear predictive coding and are very complex. However, the complexity is incorporated into digital integrated circuits and is transparent to the user. The radio industry in the United States, Europe, and Japan is gearing up for digital cellular communications and mass production of digital radios will become commonplace in the next few years.

Technical Maturity and Development Risk

Both AM and SSB are mature technologies, with low development risk. Voiceband signaling, which could be used to implement new channel access methods, channel availability improvements, and advanced user features in analog radios, is also a mature technology. Provision of such new features in an analog VHF air-ground communications system would require only engineering design, rather than research and development.

Digital radios employing modulations derived from QPSK have had significant engineering development and would represent a low-to-moderate development risk. North American digital cellular communications plans to use $\pi/4$-QPSK with differential encoding and European digital cellular uses Gaussian minimum shift keying (GMSK) [1]. These waveforms are similar to 4-OQAM and Aviation QPSK.

Microwave digital radio has made extensive use of 16-QAM modulation [2] which is very similar to 16-OQAM, so that manufacturers have experience in building radios with this modulation. Suitability of 16-OQAM for the VHF air/ground channel is less proven, however. The VHF air/ground channel experiences fast fading due to diffuse earth reflections, airframe reflections, and rotor-wing blade effects. These effects cause a rapid variation in signal amplitude and phase which may prevent the 16-OQAM demodulator from making correct decisions about which point in the signal constellation was transmitted. The critical issue is the effectiveness of amplitude and phase equalization techniques that attempt to compensate for these channel effects. Therefore, considerable analysis and experimentation is required before selecting a high data rate digital modulation like 16-OQAM for air traffic control VHF air/ground communications, where very high link reliability is required.

Backward Compatibility with AM Radios

An AM radio designed for 12.5 kHz or 10 kHz channel spacing can be made compatible with current AM radios. Two different IF bandwidths are required, a narrow bandwidth for the new channel spacing and a wider one for the 25 kHz channel spacing. An SSB radio would need two different modes, one for SSB with 6.25 kHz or 5 kHz channel spacing and

one for AM with 25 kHz channel spacing. Simultaneous AM and SSB capability at ground stations during a transition period would require two separate transmit and receive chains.

A digital radio can also have a backward compatible AM mode. The two modes will require different IF bandwidths, modulators, demodulators, AGC time constants, etc. A simultaneous AM and digital capability at ground stations will require more hardware than the either/or capability required for airborne radios. But there are fewer ground stations than airborne radios.

5.5 IMPACT/IMPORTANCE

Increasing the number of channels available for air/ground communications in the 118-137 MHz aeronautical mobile band is the most important tradeoff, because of pressures of increasing air traffic in Europe and in some parts of the United States. As we have shown, a substantial increase in the number of available channels can be provided by either analog or digital modulation. The other criteria in the first category — improvements consistent with the current concept of operations — are also important, but the demand for these improvements is less urgent. Improving power efficiency of VHF air/ground communications is desirable, but not vital. The principal benefit would be a longer communications range to high-altitude aircraft without increasing transmitter power. Reducing susceptibility to interference would result in more reliable communications and, possibly, increased frequency reuse across the country, effectively increasing the number of available channels. Some types of digital modulation employing constant envelope waveforms are better than AM in this regard, but these are not the digital modulations which maximize the number of channels per 25 kHz. There is a need for further, more detailed, tradeoff studies in this area. Better link performance in fading is a worthwhile improvement; when digital modulation is combined with error correction coding and channel equalization, it can improve link performance in fading to some extent.

Either analog or digital modulation can provide good speech intelligibility when only one talker occupies a channel — the normal operating procedure. Sometimes "walk-ons" occur, however, where two people talk at once on the same channel. Use of single sideband analog modulation can provide a degree of voice conferencing to alleviate this problem. Use of

digital modulation together with computer control of channel access can eliminate the problem.

Criteria in the second category — improvements that change the concept of operations — must be considered in the context of the evolving air traffic control system. A paper like this provides the communication engineer's perspective on what kinds of improvements are possible with new technology, but the broader perspective of the whole air traffic control community is needed to determine how operationally useful these improvements would be. This is particularly true of (1) data transmission capability as an adjunct to voice communications, (2) more flexible channel access methods, (3) improvements in channel availability and message integrity that require major changes in the ground network, and (4) advanced user features such as automated frequency changes. Compatibility with multiplexing methods and evolutionary growth capability are two other important criteria that are necessary but not sufficient conditions to accept a new modulation. Most of the criteria in the second category are better satisfied with digital modulation than with analog modulation.

The last four of the 15 criteria used to compare analog and digital modulation represent implementation issues bearing on the practicality of new modulations. As might be expected, all four criteria — need for Doppler tracking and coherent demodulation, radio design complexity, technical maturity and development risk, and backward compatibility with current AM radios — are easily met by AM and are only a little harder to meet with SSB. But the strong development effort being put forth by the radio industry for digital cellular telecommunications should make digital modulation for VHF air/ground communications practical to implement in the next few years.

5.6 CRITERIA FOR DECISION

One alternative for decision makers is to continue to use AM radios for VHF air/ground communications, but reduce the channel spacing to 12.5 kHz, or perhaps to 10 kHz, in order to obtain more voice channels in the 118-137 MHz band. A strong drawback to this alternative is that all aircraft owners must buy new, more selective, AM radios (or retrofit existing radios) while obtaining no significant benefits besides the larger number of available channels.

A second alternative is to replace AM radios with another type of analog radio; single sideband radios are the most likely choice, as they offer even more channels with a 6.25 kHz or 5 kHz channel spacing. As shown under Tradeoffs, a single sideband radio improves VHF air/ground communications in two other ways as well: (1) power efficiency is better and (2) some tolerance exists to two talkers on the same channel. However, the second alternative provides only limited benefits to aircraft owners in return for buying new radios.

A third alternative is to employ new analog radios, either AM or single sideband, both to provide more voice channels and to incorporate new features into the radios and the ground network that are of value to aircraft owners and to the air traffic control system. As shown under Tradeoffs, new features that could be provided with analog modulation are: (1) data transmission using voiceband signaling, (2) polling and trunking channel access methods, (3) path diversity and busy channel overrides using in-band signaling, and (4) some advanced user features including automated frequency handoff using in-band signaling. Providing new features with in-band signaling does require some interruption of voice, however, which may reduce user acceptance.

A fourth alternative is to employ new digital radios both to provide more channels and to incorporate new features that are of value to aircraft owners and to the air traffic control system. As shown under Tradeoffs, performance improvements and new features that could be provided with digital modulation are: (1) better power efficiency than AM, (2) some degree of performance improvement in interference and fading, (3) elimination of the "walk-on" problem, (4) data and voice transmission in the same radio, (5) more flexible channel access, (6) evolutionary growth with new vocoders, (7) flexible implementation of channel availability and message integrity improvements, and (8) automated frequency handoff and group addressing. With digital modulation, the new features can be implemented with separate TDMA signaling and control channels, so that no interference with voice channels occurs.

Before a firm choice is made among these alternatives, some technical points should be investigated in more detail, including experimentation as well as analysis. Before deciding to use analog modulation — either AM or SSB — with closer channel spacing, the number of useful channels versus the number of nominal channels needs to be determined. At ground

stations having many channels, cosited transmitters produce numerous intermodulation products and the number of intermodulation products increases faster than linear with the number of active transmitters. These intermodulation products, as well as crossmodulation products and receiver desensitization by strong interfering signals, limit the useful number of channels at airports and air traffic control centers.

Performance of digital modulation on the VHF air/ground channel needs to be evaluated by analysis, computer simulation, and field experiments. Various slow and fast fading mechanisms exist on this channel (see Tradeoffs), whose effects must be determined. Performance of channel equalization algorithms for 16-OQAM and other multiple amplitude level waveforms particularly needs to be quantified.

Performance and availability of low bit rate vocoders needs to be determined, because vocoder bit rate is a key parameter determining the number of voice channels available with digital modulation. Factors of particular interest are vocoder performance in the presence of cockpit acoustic noise, performance variability among speakers, tolerance for channel bit errors, and cost, size, and weight of devices.

Cochannel and adjacent channel interference effects on radio performance need to be analyzed and measured for the specific modulations to be employed in a new VHF air/ground radio. With analog modulation, cochannel and adjacent channel interference limit voice quality; with digital modulation they limit the usable bit rate.

5.7 CONNECTIVITY/RELATIONSHIP WITH OTHER DECISIONS

The choice between analog and digital modulation for VHF air/ground communications at branchpoint 4.1 on the decision tree is a fundamental choice that impacts the future capacity and evolution of the air traffic control system. As discussed above, the choice between analog and digital modulation relates to many other branchpoints, such as channel access (branchpoint 4.2), multiplexing (4.3.2), busy channel handling (2.1.3), workload mitigation (2.3), etc. The choice between analog and digital modulation is so interconnected with other branchpoints on the decision tree that it can only be made as part of an overall system evolution for air traffic control communications. One final thought is that this system evolution is not necessarily along a single design path, but may involve a high/low mix of

capabilities to accommodate both airlines and general aviation and to accommodate the needs of both rich and poor countries. The most likely scenario appears to be a rapid evolution towards digital radios with a wide range of capabilities for the high-end users, while an analog radio system is maintained in place for many years to accommodate the low-end users during a gradual transition to a digital capability.

SECTION 6

ANALOG MODULATIONS

6.1 CONTEXT

A systematic process for examining technical alternatives for improved air/ground communications in air traffic management has been established. A decision tree structure is shown in Appendix A that attempts to organize various alternatives in a top-down hierarchy. This provides a framework for evaluating potential solutions that can be represented by paths through the decision tree.

This paper addresses a particular alternative in the decision tree, namely, branchpoint 4.1.1, Analog Modulation. The organization of this memo is as follows: Background, Issues, Tradeoffs, Impact/Importance, Criteria for Decision, and Connectivity/Relationship with Other Decisions.

6.2 BACKGROUND

The decision tree in Appendix A considers a wide range of far-term improvements to VHF air/ground communications. Many or most of these improvements require the use of digital modulation together with computer control of access to the communications channel. In return for much increased system complexity, digital modulation and computer control can provide many new features: a greater number of communication channels, automatic handover of aircraft from one controller to another, more reliable communications, integrated voice and data, etc. Use of analog modulation in improved VHF air/ground radios is a near-term improvement of limited scope, intended primarily to address the immediate need for more voice communications channels in the aeronautical mobile frequency band. It could represent an intermediate step toward an eventual digital communications system or a lower-technology alternative for certain classes of users for whom low cost is critical.

VHF air/ground communications in the 118-137 MHz frequency band currently use amplitude modulation (AM), a technique dating from the earliest days of radio. Over the years, many other analog modulation techniques have been invented and reduced to practice,

such as frequency modulation (FM), single sideband (SSB), vestigial sideband (VSB), etc. Each of these modulation techniques has found a place in practical communication systems in applications for which their particular virtues are best suited. As a result, the advantages and disadvantages of each type of modulation are well understood.

The motivation for considering a different analog modulation than AM for VHF air/ground communications is to provide improved capability, particularly a better utilization of the frequency spectrum to make more communication channels available in the 118-137 MHz frequency band. Other improvements which could be made to VHF air/ground communications with analog modulation include:

1. Improved power efficiency; i.e., less transmitter power required to close a given air-ground link at a given audio quality.

2. Reduced susceptibility to interference and fading on the communications channel.

3. Better voice intelligibility on an air/ground link.

This decision tree paper will review the pros and cons of various analog modulations that directly modulate a radio frequency carrier with a speech waveform for voice transmission. A separate decision tree paper (section 8) will consider digital modulation techniques that first transform speech into a sequence of numbers and then transmit discrete amplitude, phase, or frequency steps of a radio frequency carrier to represent these numbers.

6.3 ISSUES

Improved air/ground communications capability can be defined by specific technical and operational criteria. The first five items below are measures of improved capability. The last two items represent costs of improved capability in terms of greater radio complexity. The criteria are:

1. Channel spacing in kilohertz, or equivalently the number of channels per 25 kHz.

2. Power efficiency, defined as the relative transmitter power required for a given audio signal-to-noise ratio.

3. Susceptibility to interference, including as a minimum co-channel interference from VHF air/ground radios in nearby geographic areas and interference due to out-of-band emissions by other services.

4. Performance degradation caused by both slow and fast fading in the communications channel.

5. Speech intelligibility, particularly in cases where two talkers inadvertently transmit at the same time on the same channel.

6. Need for doppler tracking and coherent demodulation in the radio.

7. Radio design complexity required to transmit and receive the analog modulation.

The other dimension of a comparison is, of course, the specific analog modulations to be considered. It appears that seven generic types can adequately represent the set of available modulations, as follows:

1. Amplitude modulation (AM) which lets the amplitude of a speech waveform control the amplitude of a radio-frequency carrier. AM is sometimes called double sideband transmitted carrier (DSBTC). This modulation, currently used for VHF air/ground communications, is shown in figure 6-1. The speech waveform is transmitted as both upper and lower sidebands of a radio frequency carrier.

2. Double sideband suppressed carrier (DSBSC), which transmits the speech waveform as both upper and lower sidebands, but does not transmit the radio-frequency carrier. Figure 6-2 illustrates DSBSC modulation.

3. Quadrature amplitude modulation (QAM), which transmits two non-interfering signals in the same bandwidth by modulating them onto two radio frequency

carriers with the same frequency but 90° apart in phase. As illustrated in figure 6-3, QAM can be implemented with two DSBSC signals, if the two carriers are used as phase references for modulation, but not transmitted.

4. Single sideband (SSB) which transmits a speech waveform as the upper sideband of a radio frequency carrier, suppressing the lower sideband and not transmitting the carrier. Figure 6-4 illustrates SSB, which has found wide application in high frequency (HF) radio communications and in analog frequency-division multiplexed (FDM) telephony. VHF air-ground communications has a special requirement (need to tolerate Dopper shift) which HF communications and FDM telephony do not. Modifications to SSB to accommodate this special requirement are needed, including a transmitted pilot tone.

5. Vestigial sideband (VSB) which allows part of the lower sideband to be transmitted along with all of the upper sideband. In North American television, VSB is used with a transmitted carrier for the picture signal. Figure 6-5 illustrates VSB, also including a pilot tone.

6. Frequency modulation (FM) which lets the amplitude of a speech waveform control the frequency of a radio-frequency carrier, with the carrier frequency deviation being larger than the audio frequencies in the speech waveform. As shown in figure 6-6, the transmitted frequency spectrum of FM does not resemble the frequency spectrum of the original speech. FM was invented by Edwin Armstrong in the 1920s as a means of reducing atmospheric noise (static) heard on AM radio broadcasts.

7. Narrowband frequency modulation (NBFM) which is similar to FM except that the carrier frequency deviation is reduced in order to limit the required channel bandwidth. Figure 6-7 illustrates NBFM.

To summarize, five analog modulation techniques use the amplitude of a speech waveform to control the amplitude of a transmitted signal: AM, DSBSC, QAM, SSB, and VSB. They differ in which of the two modulation sidebands are transmitted and in whether or not the carrier frequency is transmitted. Two analog modulation techniques, FM and

NBFM, use the amplitude of a speech waveform to control the frequency of a transmitted signal. They differ in the magnitude of the carrier frequency deviation.

6.4 TRADEOFFS

Table 6-1 arranges the technical and operational criteria which define improved air/ground communications capability as columns of a matrix, with the analog modulations to be considered as rows of the matrix. We will discuss tradeoffs among the modulations for each improvement criteria in turn.

Channel Spacing

VHF air/ground communications in the 118-137 MHz aeronautical mobile band currently employs 25 kHz channel spacing, but some older airborne radios are only capable of operating with 50 kHz channel spacing. Channel spacing could, in principle, be reduced to 10 or 12.5 kHz with either AM or DSBSC modulation, providing either two voice channels per 25 kHz or five voice channels per 50 kHz. A typical audio bandwidth is 3.5 kHz for telephone-quality speech. This provides a 7 kHz radio-frequency bandwidth with either AM or DSBSC (see figure 6-1). With 12.5 kHz channel spacing, this leaves 5.5 kHz between the upper sideband of one channel and the lower sideband of the next higher channel to accommodate local oscillator drift, doppler shift, and filter rolloff. With 10 kHz channel spacing, only 3 kHz is available to accommodate these factors. To realize such a large reduction in channel spacing, new radio designs will be required with closer frequency tolerances (better local oscillators or a beacon frequency tracking system), narrower IF bandwidths, and better receiver selectivity; AM radios in use today cannot operate with this smaller channel spacing. Required radio improvements are described in some detail in section 7.

Table 6-1. Comparison of Several Analog Modulations

ANALOG MODULATION TECHNIQUE	# VOICE CHANNELS PER 25 KHZ	POWER EFFICIENCY RELATIVE TO DSBSC	SUSCEPTIBILITY TO INTERFERENCE	FADING PERFORMANCE	TWO-SPEAKER CONFERENCING ABILITY	DOPPLER TRACKING REQUIREMENT	RADIO DESIGN COMPLEXITY
AM (DSBTC)	2 to 2.5	Poor -5 to -10 dB	High	Fair	Audible carrier heterodyning	None	Simple linear Tx/Rx
DSBSC	2 to 2.5	Good 0 dB	Moderate	Fair	Doppler distortion	Costas loop accurate phase	Costas loop phase tracker linear Tx/Rx
QAM	4 to 5	Good 0 dB	Moderate to High	Fair	Severe quadrature distortion	Costas loop very accurate phase	Costas loop phase tracker linear Tx/Rx
SSB	4 to 5	Good -1 dB	Moderate	Fair	Quasi-linear with different Dopplers	Pilot tone carrier error <20 Hz (voice)	Tracking loop linear Tx/Rx
VSB	3 to 4	Good -1 dB	Moderate	Fair	Quasi-linear with different Dopplers	Pilot tone accurate phase	Tracking loop linear Tx/Rx
FM	1	Good +4 dB	Low	Very good	Supression of weaker signal	None	Simple constant envelope Tx/Rx
NBFM	2	Fair -3 dB	Moderate	Good	Less supression of weaker signal	None	Simple constant envelope Tx/Rx

AM - Amplitude Modulation
DSBTC - Double Side Band Transmitted Carrier
DSBSC - Double Side Band Suppressed Carrier
QAM - Quadrature Amplitude Modulation
SSB - Single Side Band
VSB - Vestigial Side Band
FM - Frequency Modulation (25 kHz)
NBFM - Narrow Band FM (12.5 kHz)
Tx - Transmitter
Rx - Receiver

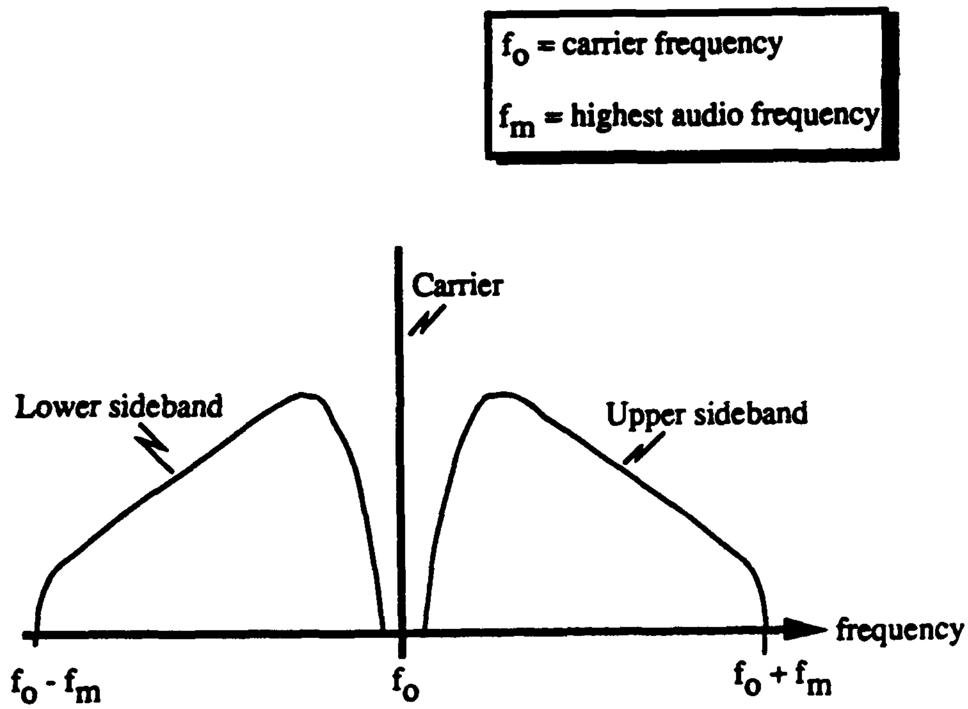


Figure 6-1. Amplitude Modulation (Double Sideband Transmitted Carrier)

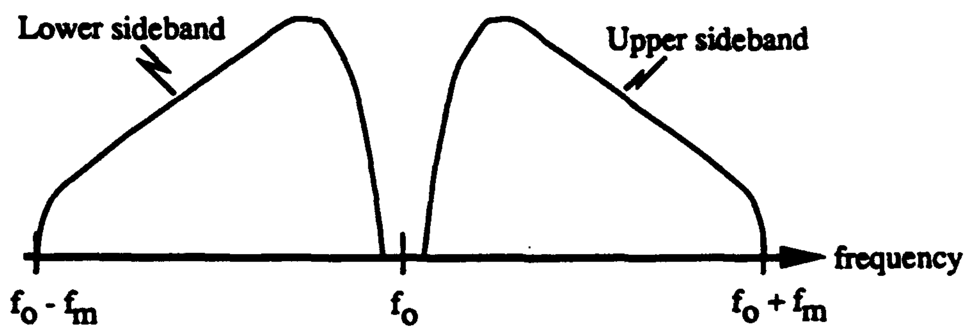


Figure 6-2. Double Sideband Suppressed Carrier

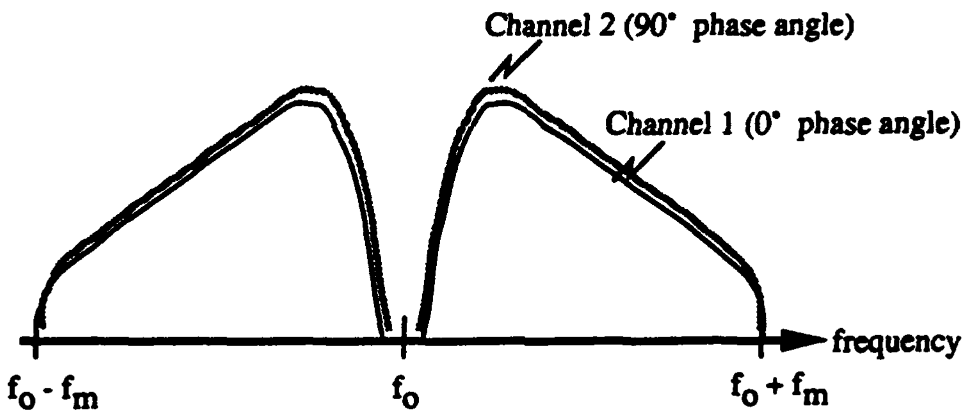


Figure 6-3. Quadrature Amplitude Modulation

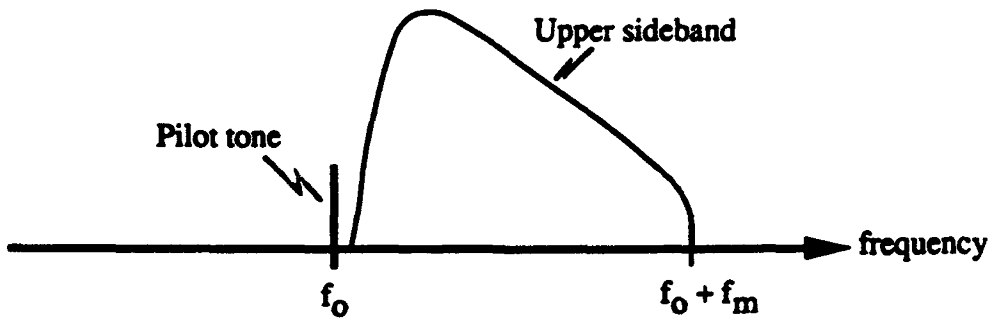


Figure 6-4. Single Sideband

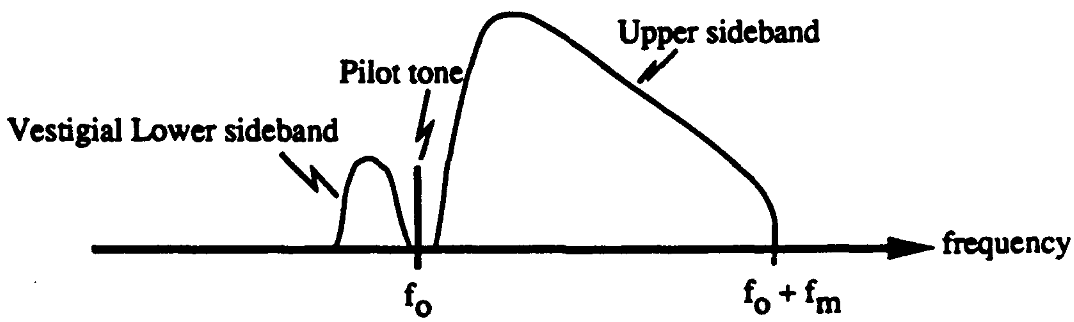


Figure 6-5. Vestigial Sideband

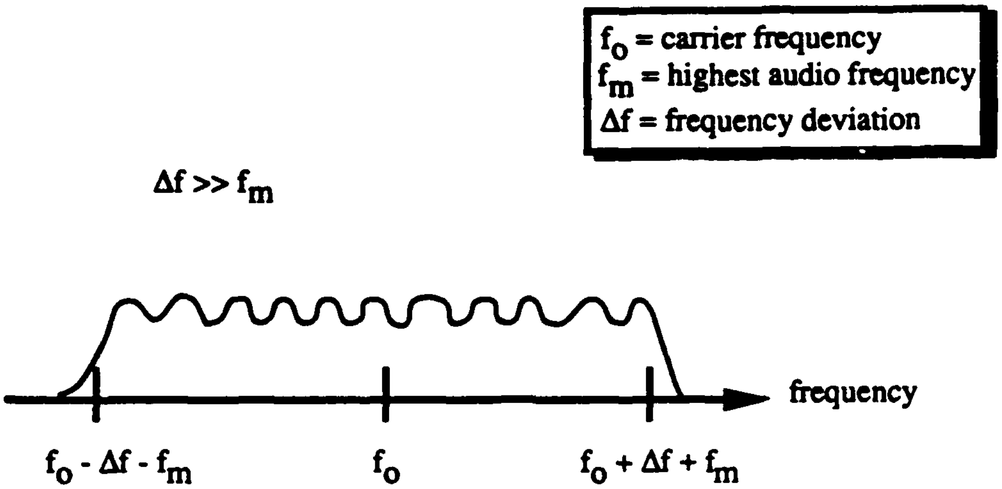


Figure 6-6. Frequency Modulation

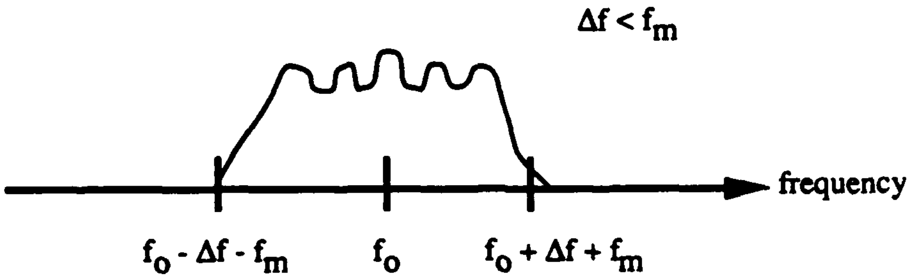


Figure 6-7. Narrowband Frequency Modulation

Because QAM multiplexes two independent signals into the same 7 kHz bandwidth, by transmitting the signals in phase quadrature, the available number of channels is double that of AM or DSBSC; i.e., four or five voice channels per 25 kHz. However, these channels must be assigned in pairs to the same ground station and the same aircraft, because of the need to transmit the paired channels in phase quadrature.

Single sideband (SSB) radios can operate with either 5 or 6.25 kHz channel spacing, providing four or five voice channels per 25 kHz. With an audio bandwidth of 3.5 kHz, the spacing between the upper end of one speech spectrum and the lower end of the next higher one in frequency is 1.5 kHz for 5 kHz channel spacing and 2.75 kHz for 6.25 kHz channel spacing. These spacings provide adequate, but not lavish, allowances for local oscillator drift, doppler shift, and filter rolloff. The large number of channels available with SSB are not subject to any pairing constraint.

Vestigial sideband (VSB) modulation is intermediate between SSB and DSBSC. It could employ a channel spacing of 6.25 kHz or 8.33 kHz, providing three or four voice channels per 25 kHz. Which of the two possible channel spacings can be used depends upon how much of the vestigial lower sideband (figure 6-5) is transmitted, which depends in turn upon the filter rolloff characteristic used to generate the VSB signal.

Frequency modulation (FM) can provide only one voice channel per 25 kHz. If one uses a frequency deviation Δf equal to 8 kHz and the highest modulating frequency f_m is 3.5 kHz, then the two sided radio frequency bandwidth 2B is given approximately by Carson's rule as $2B \cong 2(f_m+\Delta f)$, which equals 23 kHz.[1] This leaves a 2 kHz guard band between adjacent 25 kHz channels. With an 8 kHz frequency deviation, the noise quieting factor, the improvement in audio signal-to-noise ratio over the carrier-to-noise ratio, is given by $3\,(\Delta f/f_m)^2$ which equals 12 dB. This improvement in signal-to-noise ratio is the strong point of FM; fitting into narrow channel widths is not.

Narrowband frequency modulation (NBFM) could provide two voice channels per 25 kHz at the expense of noise quieting. Reducing the frequency deviation Δf to 2 kHz reduces the two-sided radio-frequency bandwidth to approximately 11 kHz, which leaves a guard band of 1.5 kHz between the resulting 12.5 kHz channels. However, the noise quieting factor $3(\Delta f/f_m)^2$ is now approximately 0 dB, so that one gives up about 12 dB in audio

signal-to-noise ratio in return for doubling the number of channels with narrowband frequency modulation, as compared to frequency modulation.

Channel bandwidth is not the only factor determining spectrum utilization efficiency. Cosite interference among multiple transmitters and receivers at a ground station (or on a large aircraft) is another limiting factor. Intermodulation product interference increases rapidly with the number of simultaneous transmissions; e.g., the number of co-site carriers present. For example, the number of two-frequency third-order intermodulation products generated by n transmitters is n(n-1). These are generally the strongest of the many intermodulation products generated by nonlinearities in transmitters, receivers, and other components. Frequency management at a ground station usually involves selecting carrier frequencies so that one does not try to receive on a frequency which is a two-frequency third-order intermodulation product of any pair of transmitters. Following this rule becomes difficult if there are many transmitters and receivers at a ground station. The overall result is that the useful number of channels at some ground stations increases more slowly than the increase in the nominal number of channels when narrower channel widths are used.

Power Efficiency

Power efficiency of AM (DSBTC) is poor because most of the transmitted power is in the carrier, rather than in the modulation sidebands that convey the speech information. Audio signal-to-noise ratio (SNR) is lower than carrier-to-noise ratio (CNR) by 5 to 10 dB, typically. The exact amount depends upon the modulation index, which constantly varies for speech. With the DSBSC or QAM modulations, no carrier is transmitted and all the transmitter power is used to send speech information. Audio SNR is 3 dB higher than the SNR at IF, on an average power basis. Since SSB reception is (conceptually) a direct translation to baseband, audio SNR is equal to the SNR at IF. At first glance, it appears that SSB is 3 dB poorer in power efficiency than DSBSC, but remember that SSB needs only half the radio-frequency bandwidth of DSBSC, so that SNR at IF is 3 dB higher for SSB for a given receiver noise level. Thus, in principle, SSB and DSBSC have the same power efficiency on an average power basis. In table 6-1, SSB is penalized by 1 dB because, in the VHF air/ground application, it needs to transmit a pilot tone for doppler tracking (discussed later) that DSB does not. Vestigial sideband (VSB) modulation is similar to SSB in power efficiency, provided the vestigial lower sideband is small. All of the suppressed-carrier

analog modulations — DSBSC, QAM, SSB and VSB — are thus seen in table 6-1 to have better power efficiency than AM because all or most their transmitter power is used for signalling and little or none is used to transmit a carrier.

Frequency modulation (FM) has good power efficiency. There is a threshold for noise improvement around 10 dB CNR (lower when threshold-extension demodulators are used). Once this is exceeded, audio SNR is about 12 dB higher than CNR, but FM needs more than three times the radio-frequency bandwidth of DSBSC. Adjusting for this, the net power efficiency advantage of FM over DSBSC is around 4 dB. Narrowband FM gives up noise improvement in the interests of a narrower channel width, as discussed above, and still has a somewhat greater radio-frequency bandwidth than DSBSC. Table 6-1 shows relative power efficiencies on an average power basis, not on a peak power basis.

Susceptibility to Interference

Comparison of analog modulations with respect to susceptibility to interference is less clear cut than comparisons of power efficiency because the results depend upon the nature of the interference. In general, AM should require a carrier-to-interference ratio, C/I, roughly comparable to the required CNR, which is in the range of 15 to 20 dB for intelligible (but not high fidelity) speech. One might expect DSBSC and QAM to require a somewhat smaller signal-to-interference ratio for satisfactory voice intelligibility, but there is limited practical experience with such modulations to draw upon. With QAM particularly, effects of interference upon the performance of the receiver phase tracking loop must be analyzed before any decision to use this modulation, since accurate phase tracking of the received signal is required to separate the two voice signals transmitted in phase quadrature. Single sideband (SSB) radios used in the HF band generally have good performance in a band which is filled with interference. However, VHF air-ground radios require a pilot tone tracking loop to remove doppler shift; HF radios do not need this feature. Thus, detailed analysis of interference effects is required before making a decision to use SSB modulation for VHF air-ground communications. Performance of vestigial sideband (VSB) modulation in interference should be similar to that of SSB.

Frequency modulation (FM) gives good performance in interference, because once the signal-to-interference ratio gets above the FM threshold, the interference is suppressed

relative to the signal by the FM detector. The resulting improvement in audio signal-to-interference ratio should be roughly comparable to the 12 dB noise improvement discussed above. Narrowband FM does not produce this improvement in audio signal-to-interference ratio at the FM detector, but its performance in interference could be better than DSBSC or SSB because it does not need to do any Doppler tracking.

Performance in Fading

VHF air/ground radio links are subject to both slow and fast fading. Slow fading is caused by such physical phenomena as specular earth reflection, foliage attenuation, atmospheric refraction, atmospheric multipath, and airborne antenna shadowing. Slow fading can be deep, with a signal loss of 10 to 20 dB or more, with the fades lasting for seconds or even minutes. All analog modulations are affected by slow fading that causes a loss in SNR at the radio receiver; this loss can be catastrophic on weak signals. With strong signals, slow fading is compensated for by receiver automatic gain control (AGC) action.

Fast fading is caused by such physical phenomena as diffuse earth reflection, particularly from buildings, airframe reflections and diffraction, and rotor-wing aircraft blade reflections. Fast fading is not as deep as slow fading, with signal loss usually less than 10 dB. The bandwidth, or doppler spread, of fast fading can be large, however: up to tens of Hertz for air-ground links in the 118-137 MHz band. If the bandwidth of fast fading extends to audible frequencies, the resulting fluctuations in signal amplitude may cause distortion of the audio for all analog modulations that are variants of amplitude modulation -- AM, DSBSC, QAM, SSB, and VSB -- even when the received SNR is high. In other words, the fast fading may be "heard" in the audio. If one makes the receiver AGC fast enough to respond to fast fading, then the AGC will cancel out lower frequency components in the audio, Thus, there is a tradeoff in the receiver AGC design between mitigating fast fading effects and preserving lower audio frequencies.

FM is much more resistant to fast fading, since signal amplitude variations do not come through the FM demodulator into the audio, provided the fading signal stays above the FM threshold. This is the reason why FM is used in the current U.S. cellular telephone system designed by Bell Telephone Laboratories [2]. The land-mobile UHF radio channel, in which the direct line-of-sight path between transmitter and receiver is frequently obstructed,

experiences deep fast fading — much deeper than the VHF air/ground channel in which a direct line-of-sight path almost always exists. The FM signal used by cellular telephones still provides good audio performance most of the time. Narrowband FM has some of the fast fading immunity of FM, but to a lesser degree because of its small frequency deviation.

Speech Intelligibility

Any of the seven analog modulations can produce good quality intelligible audio at a sufficiently high SNR. Differences in required SNR among them have already been discussed under "power efficiency." Furthermore, FM has well-known advantages over all analog modulations that vary signal amplitude, because it is immune to impulsive noise and other amplitude fluctuations on the radio channel. In order to provide a discriminant among the seven analog modulations, a special case of channel overload is considered. The discriminant among the seven analog modulations is the degree to which some speech intelligibility remains in situations where two talkers inadvertently transmit on the same channel at the same time. With AM, carrier heterodyning produces an audible tone with substantial audio distortion. If two DSBSC signals are transmitted on the same channel, substantial distortion of the audio also results because the two DSBSC signals will be offset in frequency by different doppler shifts, resulting in the phase tracking loop for coherent demodulation being mismatched to at least one of the signals. Coherent demodulation essentially "folds over" the upper and lower sidebands of DSBSC. If there is a phase or frequency tracking error, the two sidebands will not superimpose properly. With QAM, the situation is even worse, because the phase tracking errors which result from the doppler mismatch mentioned above can prevent separation of the two quadrature audio channels and cause them instead to be mixed in the audio.

A strong point of SSB is its quasi-linear response to two talkers on the same channel. With slow aircraft, the doppler shifts of the two talkers will produce pitch shifts in their audio, but some degree of intelligibility will remain. The pilot tone tracking loop which tracks signal doppler shift does not have to work perfectly with SSB to produce intelligible audio (unlike DSBSC, where any phase tracking error produces strong audio distortion). With fast aircraft, however, the pitch shifts in the audio are proportionately larger and intelligibility poorer. The pilot tone tracking loop can track only one of the two doppler-shifted pilot tones from the two talkers. Performance of VSB with two talkers should be

worse than SSB, because it has a vestige of a lower sideband and so acts somewhat like DSBSC upon coherent demodulation.

With FM, there is a strong signal capture effect. If one of the two contending signals on the same channel is significantly stronger than the other, the stronger signal will be enhanced and the weaker signal suppressed, so that the stronger signal will be intelligible. When the two contending signals are of nearly equal amplitude, they mix nonlinearly and considerable audio distortion results. With narrowband FM, the strong signal capture effect is weaker. Both contending signals mix together in the audio more frequently and produce audio distortion.

Doppler Tracking

No doppler tracking is required with AM radios. One simply makes the IF bandwidth a little wider than the two-sided signal bandwidth of 7 kHz, to accommodate both doppler shift of the received signal and local oscillator drift that together should not exceed 1 kHz at VHF. AM signal detection with an envelope detector is insensitive to this frequency shift. Accurate tracking of carrier phase is required by DSBSC to obtain an in-phase reference signal for coherent demodulation. A circuit called a Costas loop is able to recover the carrier from the symmetrical upper and lower modulation sidebands, even though the carrier is not transmitted with DSBSC [3]. If there is a phase tracking error θ, the resulting relative amplitude equals $\cos\theta$, so that θ must be held to 10 or 20 degrees. With QAM employing two DSBSC signals in phase quadrature, even more accurate phase tracking is required, since the leakage of one quadrature signal into the other channel during coherent demodulation is proportional to $\sin\theta$. Thus, phase tracking accuracies of one or two degrees are needed for good channel separation with QAM.

Single sideband is not as difficult to demodulate. A carrier frequency must be generated in the receiver, to an accuracy of about 20 Hz for voice communications. Since this accuracy is much less than doppler shift and local oscillator drift at VHF, it is necessary to transmit a pilot tone with the SSB signal and to frequency track it in the SSB receiver. Precise phase tracking of the pilot tone is not required for voice communications. Vestigial sideband modulation is between SSB and DSBSC in its doppler tracking requirements, since part of a lower sideband is transmitted.

No doppler tracking is required by FM or NBFM. A limiter-discriminator FM detector converts the frequency modulated IF signal to an amplitude modulated audio signal. Doppler shifts at IF cause DC level shifts at audio which the ear cannot detect.

Radio Design Complexity

An AM radio is simple to design and manufacture. It does require a linear power amplifier to transmit the peaks of speech signals and a linear receiver with AGC to receive the AM signal. A DSBSC radio is much more complex. Besides circuitry to suppress the carrier in the transmitter, a coherent receiver is required for synchronous demodulation, including a Costas loop to recover and phase track the (suppressed) carrier. A linear transmitter and receiver are also required.

Some of the circuitry in a DSBSC radio must be duplicated to transmit and receive QAM. Two audio signals must each be DSBSC modulated, added together in phase quadrature, and then amplified in a linear power amplifier, so as to produce no intermixing of the two superimposed channels. In the receiver, a Costas loop is needed to recover and phase track the (suppressed) carrier. Because of the channel pairing on platforms, the two quadrature channels have the same doppler shift and local oscillator frequency drift, so that only a single phase tracking loop is required. Two synchronous demodulators are required to recover the two superimposed DSBSC signal channels. Accurate phase tracking and circuit linearity are required to obtain good separation of the two quadrature channels.

An SSB transmitter requires a linear power amplifier. Linearity requirements for SSB are more stringent than those for AM or DSBSC, because of the higher peak-to-average ratio of an SSB signal. Automatic level control (ALC) design for the power amplifier is more difficult than for AM; unlike the AM carrier, the SSB pilot tone may not have sufficient amplitude to drive an ALC. An SSB receiver needs a pilot tone tracking loop to remove the local oscillator offset between transmitter and receiver as well as the doppler shift and, furthermore, the performance of this frequency tracking loop is critical to audio quality and signal intelligibility. Tradeoffs exist between acquisition time of the tracking loop and frequency tracking accuracy. The receiver RF, IF, and audio amplification chain must be

linear to accommodate signal peaks. Transmitter and receiver complexity for VSB are comparable to that for SSB, but pilot tone tracking accuracy must be better.

An FM transmitter is simpler than an AM transmitter because of the constant envelope waveform used for FM. An FM receiver does not need to be linear, but an FM discriminator is more complex than an AM envelope detector. No AGC is required in an FM receiver, but filter group delay can cause audio distortion, requiring careful design of filters. Narrowband FM differs from FM only in the size of the frequency deviation employed. Radio design considerations are similar for both.

6.5 IMPACT/IMPORTANCE

Increasing the number of channels available for air/ground communications in the 118-137 MHz aeronautical mobile band is the most important tradeoff, because of pressures of increasing air traffic in Europe particularly. There are differences among the alternative analog modulations: at least a doubling of number of channels is possible in principle with AM, DSBSC, or NBFM; approximately a quadrupling of number of channels is possible in principle with QAM, SSB, or VSB; but no improvement is possible with FM. One must note, however, that the number of channels which can be used at a single ground station or in a small geographical area does not increase as fast as the number of available channels, because of the cosite interference effects discussed earlier. Thus, in many areas where air traffic density is high, the increase in the number of useful channels obtained by closer channel spacing with analog modulation is significantly less than the apparent increase in total number of channels.

Improving the power efficiency of VHF air-ground communications is a desirable, but not vital, improvement. Changing from AM to any other analog modulation improves power efficiency. As shown in Table 6-1, FM gives the greatest improvement in power efficiency but, as just noted, FM gives no improvement in channel spacing, a factor more important than power efficiency.

Susceptibility to interference is more important than power efficiency, but there are not large differences among the various analog modulations. FM should have the best performance, but has the drawbacks already noted. Single sideband has promise of better

performance than AM, based upon experience in the HF band, but the special requirements for Doppler tracking in aeronautical mobile communications must be considered.

Resistance to fading is of some importance, although the performance of AM is currently at least satisfactory. Changing to another form of amplitude modulation like DSBSC, QAM, SSB, or VSB may not produce a large improvement in resistance to fast fading. Either FM or narrowband FM would give an increased immunity to fast fading.

Voice quality and intelligibility are important; any of the analog modulations are satisfactory at a sufficiently high signal-to-noise ratio. Retaining speech intelligibility with two talkers inadvertently using the same channel may or may not be important depending upon operational procedures. If it is important, then SSB and FM have significant advantages over the other analog modulations. Some degree of two-talker conferencing is provided by SSB, while FM allows the stronger of two contending signals to suppress the weaker and become intelligible.

Need for doppler tracking and radio design complexity are both important factors in choosing an analog modulation. The accurate phase tracking and circuit linearity required by DSBSC and QAM, coupled with the radio industry's limited experience in constructing radios using these modulations, weigh against their choice. Although an SSB radio is more complex than an AM or FM radio, the radio industry has a lot of experience building SSB radios and has learned how to deal with this complexity. The easiest improvements to implement would be AM or narrowband FM with 12.5 kHz channel spacing, which lead to simple radios with noncoherent demodulators and no requirement for doppler tracking.

6.6 CRITERIA FOR DECISION

The criteria discussed at some length above are the number of voice channels provided by each analog modulation, relative power efficiency, susceptibility to interference, performance in signal fading, intelligibility under a two-speaker overload condition, need for Doppler tracking, and radio design complexity. Significant differences are found to exist between analog modulation types when compared against these criteria. There are, in addition, a number of other criteria which must be considered in making a decision on which modulation to use for a future VHF air/ground communication systems. These include life

cycle cost, technology trends, the transition from the present system to a new system, the need to send both voice and digital data in the same radio, etc. Against the criteria used for a detailed comparison, some clear trends appear.

Analog modulation allows in principle either an approximate doubling or a quadrupling of the number of channels in the 118-137 MHz frequency band. The number of channels can be doubled with AM, DSBSC, or NBFM. There are no compelling reasons to choose DSBSC over AM, except for power efficiency which is not one of the more important decision criteria; together with the much greater complexity of a DSBSC radio over an AM radio, this tends to rule out DSBSC modulation as a choice. Both AM and narrowband FM are simple to implement and well understood by industry; the choice between them could be deferred until a finer-grain comparison is made. One aspect of such a comparison would be to calculate the increase in the useful number of channels in situations where cosite interference is a limiting factor.

If an approximate quadrupling of the number of channels is required, this can be done in principle with either QAM or SSB. There are no compelling advantages of QAM over SSB, while QAM is more complex to implement than SSB, its performance suffers much more from phase tracking errors, and its channels must be paired on the same platform. Thus, SSB appears to be the analog modulation of choice for a large increase in number of channels. This large apparent increase in the number of channels must be tempered with practical limitations of cosite interference in some geographical areas. There is no good reason to chose VSB over SSB. The strong point of VSB is a baseband frequency response extending to very low frequencies. While this factor is important to reception of television pictures (for which VSB is presently used) it is not important to voice transmission.

Finally, although FM with a large modulation index has a number of good points -- simple implementation, very good power efficiency, resistance to fast fading, and interference rejection -- it cannot provide more channels and can probably be dropped from consideration.

6.7 CONNECTIVITY/RELATIONSHIP WITH OTHER DECISIONS

This paper discusses the choices among analog modulations, branchpoint 4.1.1 on the decision tree. A more basic decision, at branchpoint 4.1, is whether to employ analog or digital modulation, since digital modulation can also provide an increased number of voice channels in the 118-137 MHz band, better power efficiency, and a greater degree of resistance to interference and fading than the presently used AM radios. It does this at the cost of much more complex radio design. Design of the radio-frequency and intermediate-frequency stages of a digital radio requires careful consideration of group delay effects on digital symbol transmission. Considerable baseband signal processing is required to encode speech digitally at low data rates and to provide multiple channels with time division multiple access. A critical consideration in choosing between analog and digital modulation is whether or not voice and data communications are to be provided in the same radio. If they are, then the decision will probably be for digital modulation for both functions, because of the incompatibility of analog radio design with narrow channel spacing for voice with the needs of data transmission. Narrowband filters optimized for voice signals have group delay characteristics which cause intersymbol interference in data transmission. On the other hand, if a separate voice radio is to be retained, then analog modulation probably leads to a lower cost radio than digital modulation to provide a given number of channels.

Finally, many of the branches in the decision tree relating to channel access (under branchpoint 4.2) are more compatible with digital modulation than with analog modulation. Essentially, analog modulation is a simpler and nearer-term improvement which addresses primarily the need for more channels, while digital modulation is a more complex longer-term improvement which provides a number of other features besides a greater number of channels.

SECTION 7

SPECTRUM UTILIZATION - CLOSER CHANNEL SPACING

7.1 CONTEXT

A systematic process for examining technical alternatives for improved air/ground communications in air traffic management has been established. A decision tree structure is shown in Appendix A that attempts to organize various alternatives in a top-down hierarchy. This provides a framework for evaluating potential solutions that can be represented by paths through the decision tree.

This paper addresses a particular alternative in the decision tree, namely, branchpoint 6.1, Closer Channel Spacing. The organization of this paper is as follows: Background, Issues, Tradeoffs, Impact/Importance, Transition, Criteria for Decision, and Connectivity/Relationship with other Decisions.

7.2 BACKGROUND

The decision tree considers both near and far term improvements to VHF air/ground communications. Many of the suggested architectures offer increased user capacities and spectrum efficiency through the use of digital modulation techniques and multiple access schemes. Analog modulation techniques have evolved since the invention of double-sideband transmitted carrier amplitude modulation (DSBTC - AM) which is still found in the voice communication systems of not only the ATC industry, but also in military voice communications. Suppressed carrier single sideband (SSB) and vestigial sideband (VSB) modulation techniques offer increased bandwidth efficiencies at the expense of circuit complexity and cost. While many advanced analog modulation techniques offer a two to three fold increase in the number of available channels as compared to AM, digital modulations offer even higher channel capacities through the use of multiple access architectures[1]. In order to increase the number of channels within a limited frequency allocation such as the ATC VHF band, 118 to 137 MHz, we must consider for each of the

[1] Additional information on analog modulation, digital modulation, and multiplexing can be found in sections 6, 8, and 10, respectively.

candidate modulation techniques, the design requirements of their use, and the overall system impacts during transitional periods.

Several system components and parameters in a transceiver are influenced by the choice of modulation: IF filter, RF and IF gain characteristics, instantaneous dynamic range, synthesizer tuning time, and transmitter characteristics. Each will be discussed relative to the various proposed modulation schemes and channelization.

7.3 ISSUES

The current 25 kHz channel allocated for conventional VHF and UHF AM voice communications stems from the technology available during the mid 1970's for reference oscillators, crystal filters, and solid state power devices. The improved performance of these devices and the rapid evolution of solid state devices allowed for the transition to 25 kHz from the original channelization of 50 kHz during the post WWII years. This same evolution now offers the designer many options to increase the system information rate and spectral efficiency. This must be levied against the user requirements of the future and the projected cost.

Airborne transceivers are burdened with small volumes, large temperature swings, noisy supply voltages, and severe vibration profiles as compared to their ground counterparts. The current VHF ATC band covers 118 MHz to 137 MHz, which can be divided into the following number of channels depending on the selected channel bandwidth,

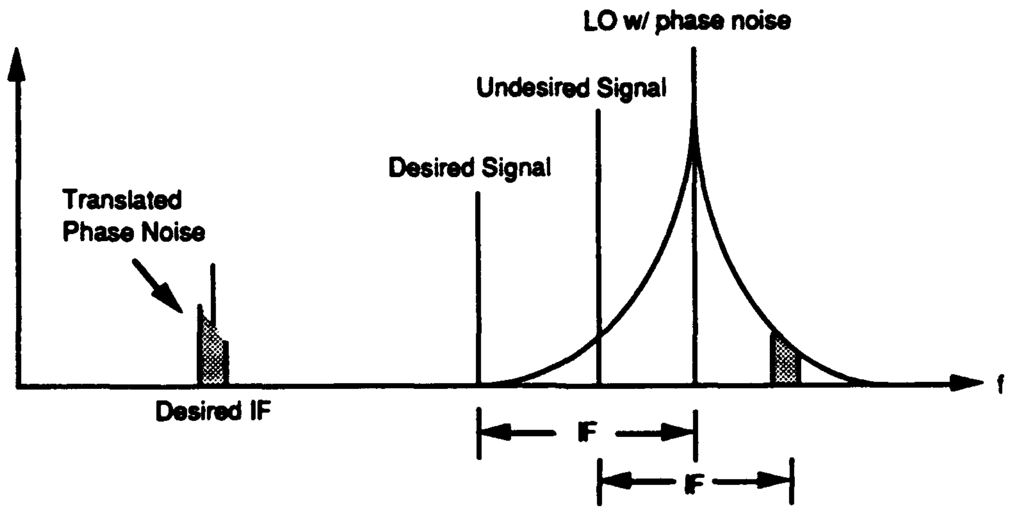
6.25 kHz	3040
12.50 kHz	1520
25.00 kHz	760
50.00 kHz	380

The implementation of narrow frequency assignments affects both the design requirements of a new radio and compatibility with the existing systems.

Frequency synthesizer

One obvious impact on the transceiver design is the synthesizer frequency resolution and stability. Most single channel analog radios of today utilize single chip digital Phase Lock Loop (PLL) integrated circuits (ICs) in a second or third order PLL synthesizer. The low-end general aviation AM transceivers will have 25 kHz synthesizer step sizes, while the commercial aviation transceivers can have resolutions as small as 5 kHz to support multi-carrier voice and UHF navigation functions. It is not uncommon to have integrated voice and navigation functions within the same unit.

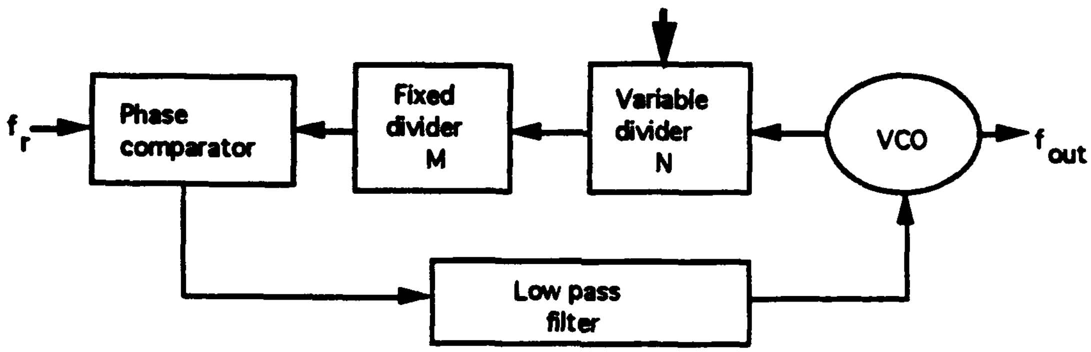
The increase in the PLL frequency resolution has a significant impact on the synthesizer design in terms of attaining the step size while providing sufficient adjacent channel rejection. The synthesizer performance is governed by the adjacent channel rejection and spurious emission requirements of the system. For example the noise pedestal about the LO can effectively translate undesired energy from a nearby strong channel into the same IF. The figure below illustrates this effect.



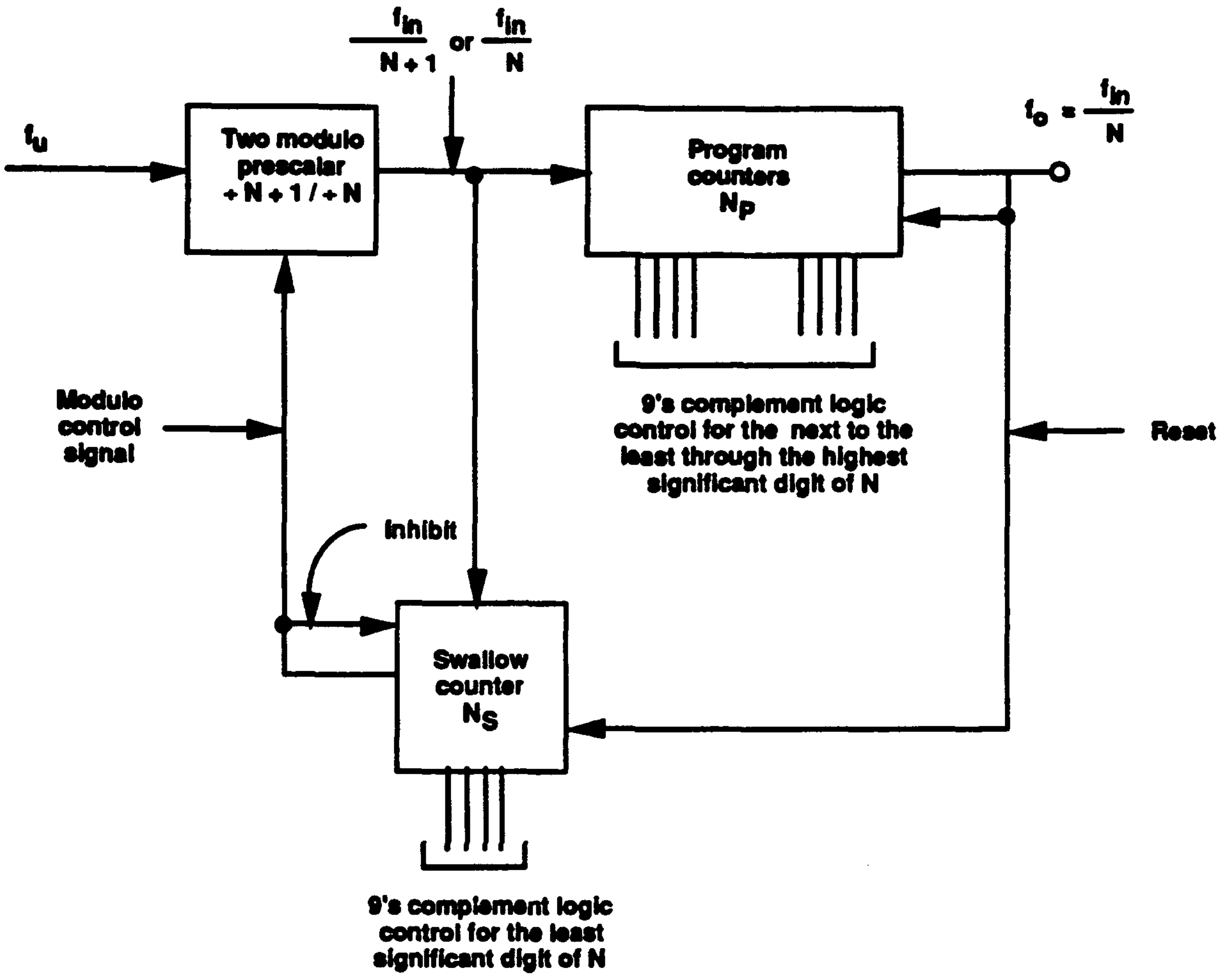
If the effective adjacent channel rejection (suppression of the adjacent channel) is specified to be 70 dB for example, then the SSB noise level within the receiver bandwidth must be lower than the 70 dB attenuation point. The performance gained by a high order, high quality factor (Q), narrow bandwidth IF filter could be completely negated by the phase

noise characteristics of a poorly designed PLL or of a PLL designed for a system with a wider channel separation.

The design of the PLL with small frequency steps requires a number of design tradeoffs. The basic form of a digital PLL is shown in the figure below. The loop consists of a Voltage Controlled Oscillator (VCO), a fixed frequency divide by M, a variable frequency divide by N, a phase comparator, and a low pass filter.



As suggested by the basic block diagram the output frequency step size is M times the reference frequency, where M is the fixed prescaler ratio. In a fine frequency stepped system where every channel could effectively be used, the reference frequency must be reduced by this same factor M. As a result the loop bandwidth must be reduced to maintain loop stability. A conflict arises because the loop bandwidth also determines the SSB phase noise characteristics and the loop settling time. In general the loop bandwidth is made as wide as possible to minimize settling time and offer the greatest degree of shock and vibration immunity. The loop bandwidth however, can not be greater than the reference frequency. The fixed divider ratio is the limitation in this form of a PLL. One alternate solution is the two-modulus prescaler or pulse swallowing approach where a prescaler is selectable to two integer values, N or N+1, followed by two programmable dividers operated in parallel. This method can allow for a frequency reference equal to the channel spacing. A block diagram is shown below.



The general characteristic of the PLL is that it acts as a low pass filter for signals inside the loop bandwidth and as a high pass filter for signals outside the loop bandwidth. Phase noise from the oscillator inside the loop bandwidth will be removed while components outside the loop will not be affected. Frequency instabilities in the synthesizer output are directly related to the instabilities of the reference oscillator. The phase noise characteristics are simply the frequency domain transforms of the reference oscillator instabilities. The synthesizer output phase noise bandwidth is related to the reference oscillator phase noise bandwidth multiplied by the divider ratio.

In summary the choice of loop bandwidth affects close-in phase noise and lockup time. Phase noise is produced by the dividers, phase detectors, and filters, and when multiplication ratios are high, the reference frequency phase noise can be dominant when multiplied. To

minimize this effect, the loop bandwidth can be narrowed since the majority of the noise outside the loop is solely determined by the VCO. Typical dividers have phase noise characteristics of -150 to -160 dBc/Hz which places most of the phase noise characteristics inside the loop to be dominated by the reference oscillator. The reference oscillator for low cost PLLs is usually a crystal controlled oscillator which offer stabilities of 30 to 50 ppm while higher end PLLs will have a Temperature Compensated Crystal Oscillator (TCXO) which provides stabilities of 1 to 3 ppm (for -20 C to +70 C). The costs associated with a synthesizer capable of supporting a system with 12.5 kHz channels are dominated by the reference oscillator which is discussed below.

Reference Oscillator

As mentioned above the phase noise performance can be strongly influenced by the choice of reference oscillator. The performance of the oscillator is a function of the stability and emission requirements of the system, the adjacent channel rejection requirement, and the specified environment. Most commercial avionics are specified to adhere to RTCA DO160B for temperature and altitude limits. The typical requirement is category A1 for general aviation and D1 for commercial aviation. The details of each category are outlined in the table below.

	A1	D1
Low Operating Temp	-15 C	-20 C
High operating Temp	+55 C	+55 C
High Short -Term Operating	+70 C	+70 C
Low Temperature Ground Survival	-55 C	-55 C
High Temperature Ground Survival	+85 C	+85 C
Altitude	15 kft	50 kft

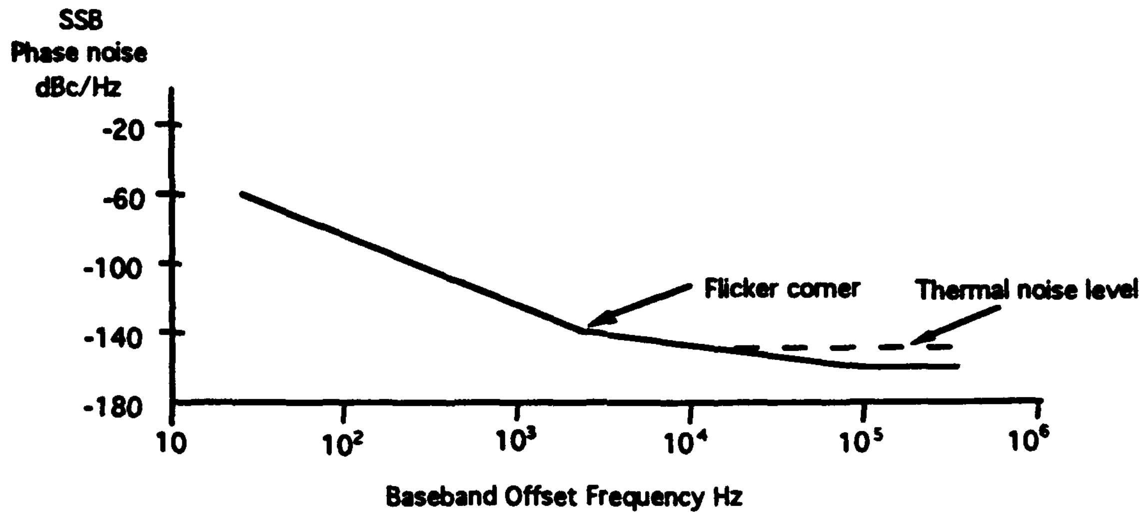
Most commercial airborne transceivers are designed to operate within -20 C to +55 C while the ground transceivers are specified to operate within -30 C to +60 C. The typical ground transceiver is specified at a frequency stability of 10 ppm after a five minute warm-up. The stability performance of the airborne transceiver depends on the cost and features of the specific unit. Low-end radios will have stabilities of 50 ppm or higher, while the stability of dual communications/navigation (COM/NAV) radios will be approximately 20 to 30 ppm.

The cost of the reference oscillator is inversely proportional to its stability. Crystal and crystal oscillator manufacturers were contacted (March 1991) to obtain estimates of the costs for three of the most commonly used types of reference oscillators. Costs can vary significantly depending on the popularity of the desired oscillator frequency and the volume needed. The cost of each category of reference oscillator is provided in the table below.

Type	Nominal Stability	Cost (quantity- 100)
Discrete Crystal Oscillator	50 ppm	$ 1
Packaged Crystal Oscillator	10 ppm	$ 20
TCXO	1 - 3 ppm	$175

Voltage Controlled Oscillators (VCOs)

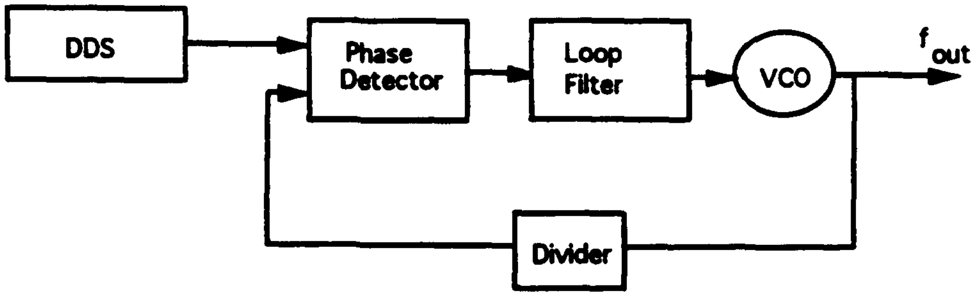
In a single loop synthesizer the phase noise performance outside the loop bandwidth is dominated by the VCO. The varactor controlled oscillator dominates most designs in this frequency range. The SSB phase characteristics of VCOs depends on the oscillator design, specifically the transistor parameters and overall resonator Q over the desired bandwidth. The SSB phase noise of a VCO is typically discussed in terms of the contributing elements in the oscillator design. The basic oscillator is formed by cascading an amplifier with a resonator then closing the loop. The resonant sections in a VCO are tunable via mechanical or active means. The extremely close-in noise contribution is due to amplifier *flicker* noise. The oscillator amplifier and any buffer amplifier outside the resonant loop contribute to the flicker noise. The offset frequency at which the flicker noise is equal to the thermal noise level is referred to as the flicker corner. The rolloff characteristic of the SSB phase noise as a function of the baseband offset frequency is illustrated below.



The second dominant source of phase noise is the modulation phase noise. Due to the presence of tuning elements within the resonator, any internal or external noise voltages present on the tuning element will offset the desired frequency proportional to the VCO gain constant. The desired tuning bandwidth of the VCO and the available supply voltage are traded-off against the modulation noise in designing the VCO. The lower the VCO gain constant, the lower the modulation noise contributions to the overall SSB noise. The majority of the VCOs within commercial VHF radios are discrete designs to minimize material costs, although many manufacturers offer small hybrid packages with excellent performance. Discrete VCO designs which require low phase noise and have small tuning bandwidths are easily achievable for a few dollars in this frequency band.

Digital Direct Synthesis

One method of frequency generation which has replaced the PLL in certain applications is the Digital Direct Synthesizer (DDS). This device, when used as a frequency synthesizer, produces a sinusoidal output derived from a digital representation which has been converted to analog with a high-speed digital-to-analog converter (DAC). One commonly implemented configuration of a low cost, small volume, high speed synthesizer is a hybrid DDS/PLL. In this form the DDS is used as the reference of a coarse tuned PLL. The fine frequency step size and speed of the DDS are utilized outside the loop bandwidth. A block diagram is shown below.



The hybrid DDS/PLL approach is very versatile in applications where fast switching times and fine frequency resolution are needed. The fine frequency and phase resolution of the DDS allows for direct programming of complex digital modulations and quadrature IF processing. Since the majority of the circuitry in the DDS is digital, extremely small sizes are possible with VLSI circuitry. Multi-loop high order PLL designs of comparable performance typically will occupy larger volumes and be more costly. The downside of DDS technology is the close-in phase noise characteristics. The phase noise characteristics of a hybrid DDS/PLL synthesizer available with today's technology will be on the order of -65 dBc. The spurious performance of the DDS has been predominantly limited by the performance of the DACs. However, DAC performance has improved significantly throughout the evolution of DDS technology where high performance DDS synthesizers of today can achieve spurious levels below -70 dBc. A finely channelized transceiver would require a synthesizer with phase noise levels of -70 to -80 dBc *at the adjacent channel* (± 12.5 kHz). Hybrid PLL/DDS or DDS synthesizers of this performance level cost approximately $100 (1992 $) in small quantities which is at least an order of magnitude higher than the present day PLL synthesizers.

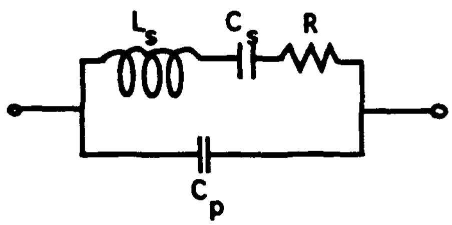
A transceiver designed for 25 kHz channelization will have design tolerances related to adequate signal detection within this bandwidth for the specified environment. The oscillator stability, synthesizer phase characteristics, and the synthesizer step size would need to be improved via a redesign to upgrade an existing transceiver to even 12.5 kHz from 25 kHz channels. In the design of a new system based on 12.5 kHz channels, the synthesizer and reference oscillator would need to have the respective phase noise and stability characteristics to achieve the benefits of the narrower channels.

IF Filters

The characteristics of the final IF bandpass filter enhances the selectivity and provides the channel-to-channel isolation of the transceiver. The final IF in a typical triple conversion superheterodyne VHF receiver can reside at frequencies ranging from 455 kHz to 21.4 MHz. The final predetection IF frequency is critical to the design of the final filter, since the percent bandwidth decreases proportionally with frequency. The percent bandwidth dictates the Q needed to achieve the desired passband characteristics. When the desired bandwidths of bandpass filters becomes narrow, lumped element filters become impractical due to the limitation in the resonator's Q. In a lumped element filter the unloaded Q can be represented by the relationship below.

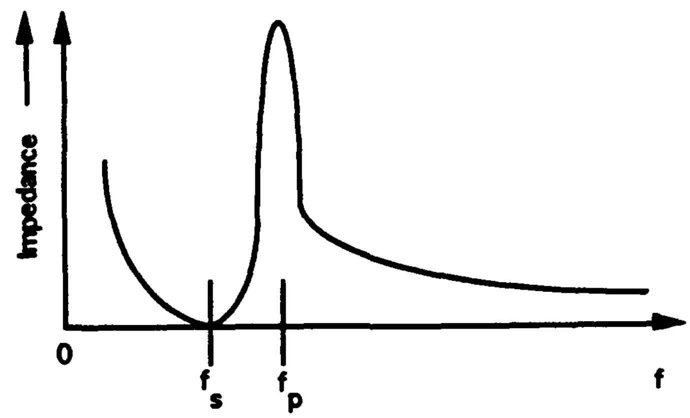
$$Q = q_0 \frac{f_c}{BW_{3\ dB}}$$

The value of q_0, the realizable Q of each individual lumped element component, is limited by the resistance of the inductive elements in each resonant section. At HF and VHF frequencies the minimum Q of inductors ranges from 10 to 400, primarily limited by high resistances. For this reason crystal filters dominant most narrow IF filter applications. The losses in a crystal which forms a portion of each resonant section are dominated by internal dissipation, mechanical mounting, and motional damping. All of these are negligible as compared to lumped elements. Qs can range from 10^5 to over 10^6. The crystal can be utilized by understanding the equivalent circuit of a crystal as shown below,



The crystal supports both parallel and series resonances where each can be used to create unique filter responses with extremely high Qs. The typical response for a single

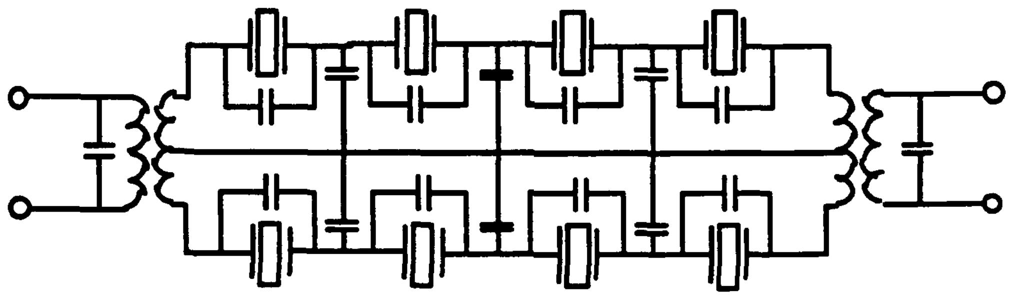
excited crystal is further illustrated by the diagram below depicting impedance versus frequency,



where f_s and f_p, respectively, are the series and parallel resonant frequencies corresponding to the equivalent circuit.

The ability to create high values of attentuation in the immediate stopbands is a major factor in designing narrow band IF filters where adjacent channel rejection is a significant system requirement. The basic pole-zero response shown in the previous diagram, combined with additional lumped elements, can create bandpass filters of moderate order with regions of high attenuation in the close-in stopbands. Other characteristics must also be considered when choosing a predetection filter, such as: ripple, phase, and group delay. The specification of each of these parameters depends on the type of modulation present on the waveform. Non-linear phase and group delay characteristics can severely degrade both the time and frequency domain characteristics of the desired signal prior to final detection. For example the parameters of flat group delay and sharp skirt response are usually opposing. Recalling that the Hilbert Transform of an infinite attenuation response is an infinite time delay response, sharp narrow bandwidth filters will have non-uniform group delays across the passband with high values of delay at the filter corner or cutoff frequecies. However, in a communication system designed on 12.5 kHz or less centers, the filter skirts and immediate stopbands must offer values of attenuation of approximately 80 to 90 dB. This can result in possible waveform distortion due to non-linear phase or non-uniform group delay. To further illustrate this issue, a standard 10.7 MHz bandpass crystal filter that could provide the necessary characteristics for a 12.5 kHz system was modeled using a RF CAD routine. The

structure is a eight-pole bandpass filter optimized for at least 80 dB rejection in the adjacent 12.5 kHz channels and providing linear phase within the passband. The schematic is shown below followed by the results of the simulation. The results will be displayed in terms of the insertion loss (S21), the return loss (S11), the group delay, and the phase response.



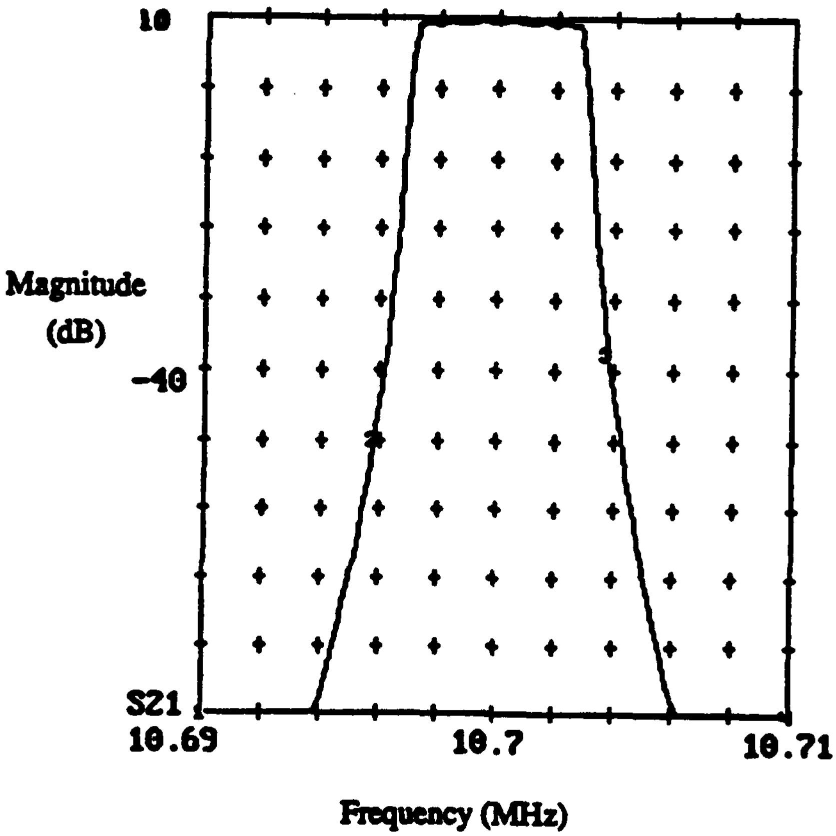


Figure 7-1. Insertion Loss (S21)

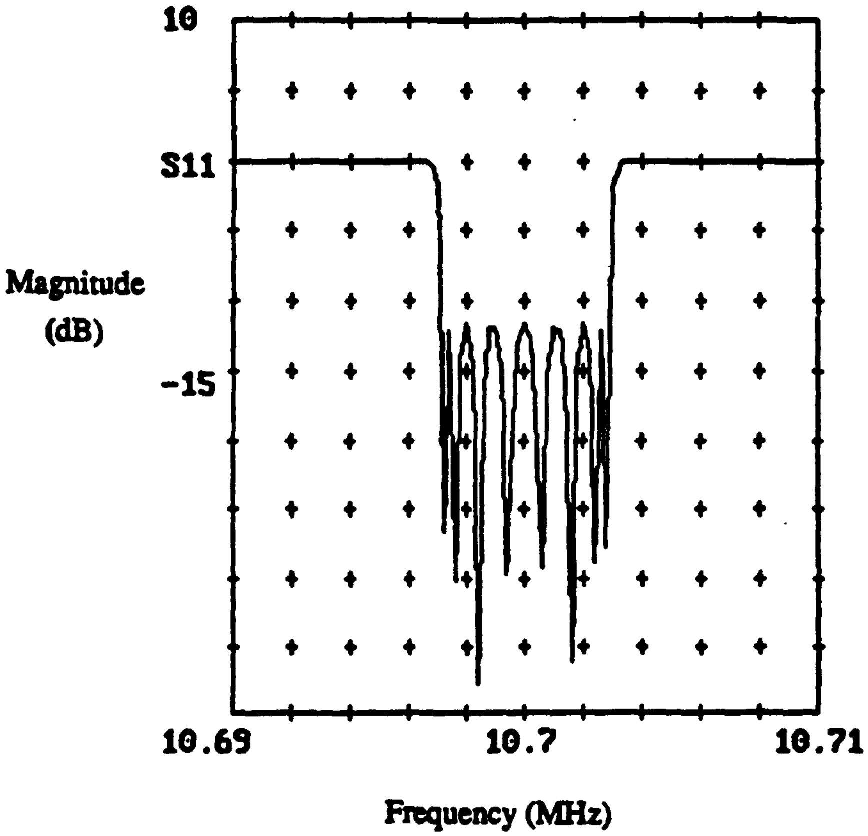


Figure 7-2. Return Loss (S11)

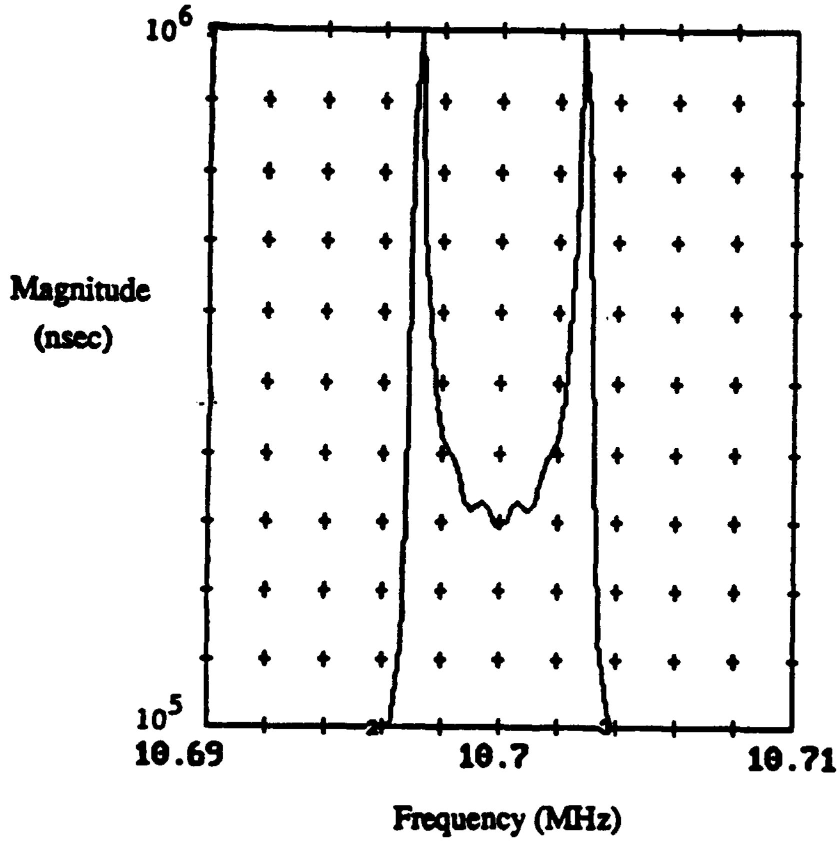


Figure 7-3. Group Delay

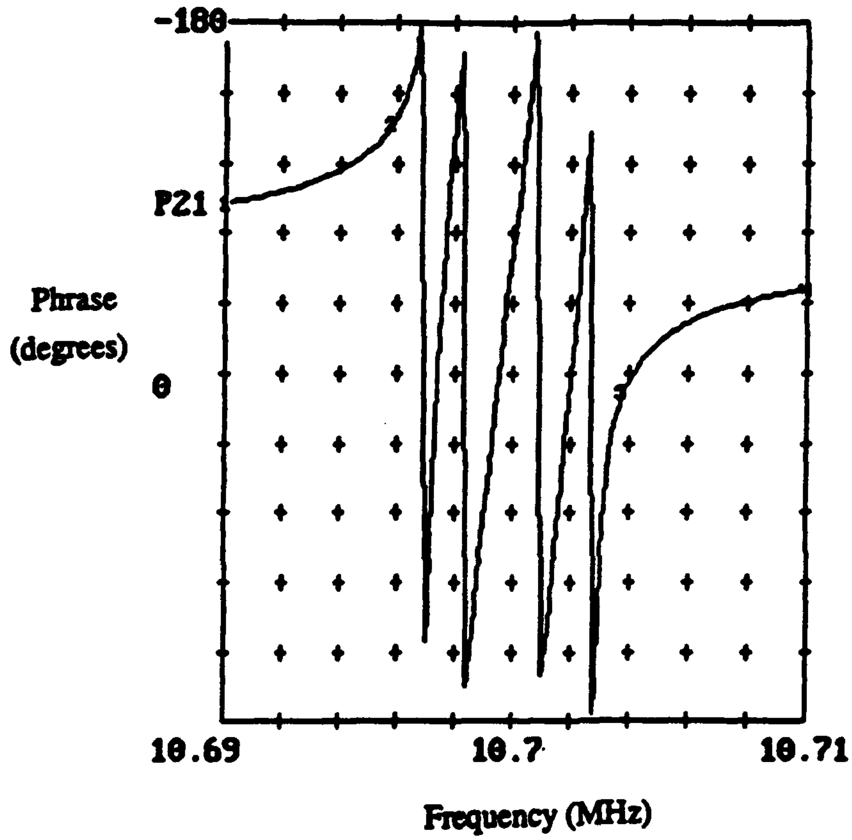


Figure 7-4. Phase Response

Note that in the insertion loss graph the adjacent channel attenuation is greater than 80 dB with a 3 dB bandwidth of 6 kHz in the passband. The return loss graph depicts the input impedance match conditions,while illustrating the placement of each pole in the passband. The group delay response illustrates the magnitude of variation possible across the passband of a narrow, high order filter, while the phase response shows a linear response within the 3 dB bandwidth. Group delays of this magnitude, if left uncorrected in pre- or post-demodulation equalization, would result in severe intersymbol interference for most frequency varied modulations. Although a filter with the above characteristics is obtainable from several vendors at moderate cost (within the same order of magnitude of the lower order filters within the transceivers of today), the pre- or post-demodulation processing can require a significant increase in componentry if performed with discrete circuits or require a microprocessor if done digitally. This departs from the simple audio circuits which are present in the radios of today.

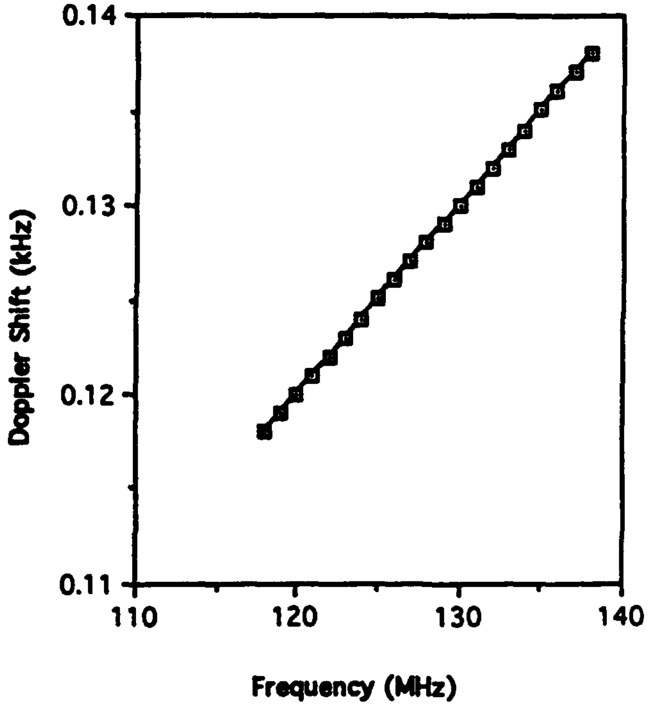
Modulation techniques which utilize multiple phase states and/or frequency shifted waveforms will require linear phase and uniform group delay responses. Alternate filter designs could be shown which are optimized for equi-delay while preserving a certain degree of the linear phase or stop-band rejection characteristics. Although such structures are realizable, they are considerably more expensive. The significant issue of the IF filter choice is specifically related to the type of modulation used in the system and the system performance requirements.

7.4 TRADEOFFS

In the design of a system with 12.5 kHz channels, the channel bandwidth must be divided into two basic contributing elements: first, the required information bandwidth for error free communication and secondly, the remainder for carrier offsets due to frequency stability and Doppler. A review of each element would justify approximately 6 kHz for the information bandwidth, 0.140 kHz for Doppler offset, and the remainder for synthesizer drift at both ends of the link. The performance tradeoffs in the design of each system component related to this allocation must be levied against unit cost and user requirements. The tradeoffs are specifically discussed below by applying the above technological design issues to the allocated frequency budget.

Doppler Shift

Commercial aircraft of the 1990's approach cruise velocities of approximately Mach 1.0. The maximum Doppler shift for a fixed (ground) transmitter and an airborne receiver will be the aircraft velocity over the operational wavelength of the received signal. This is illustrated in the figure below. The Doppler shift will be up to twice this amount for air-to-air links.



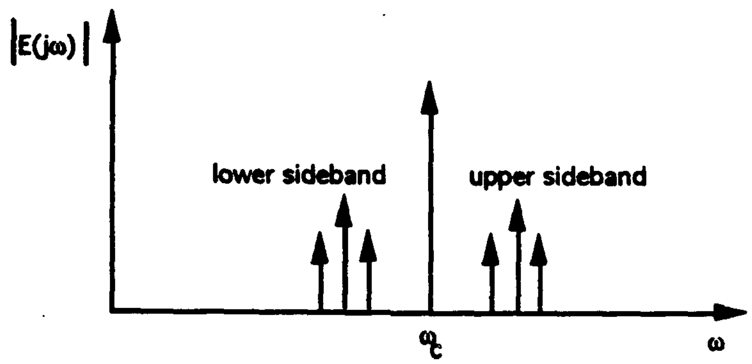
Spectral Containment - Information Bandwidth

The frequency spectra of the transmitted waveforms for conventional amplitude modulation will vary depending on the purity of the power supply circuitry, the linearity of the RF amplification stages, and the purity of the synthesizer. The time domain

representation of the modulated carrier, with the highest baseband frequency of order m, can be expressed as

$$e(t) = A_0\left[1 + A_0^{-1} \sum_{k=1}^{m} A_k \cos(\omega_k t + \theta_k)\right]\cos(\omega_c t)$$

which can be spectrally illustrated by the figure below.



The modulation sidebands evident in the spectral display above are the primary frequency products resulting from conventional amplitude modulation. Conventional AM is also referred to as double-sideband transmitted carrier amplitude modulation (DSBTC-AM). The spectral components in *each* sideband contains both the amplitude and phase information of the original baseband signal. The carrier itself contains no information. Transmission of the entire spectrum represents an inefficient use of the frequency spectrum and the system power.

SSB systems exist which transmit only one of the sidebands while suppressing the carrier and the alternate sideband. Spectral efficiency is achieved since half the RF bandwidth is needed for transmission and the RF output is only present when a baseband signal is applied to the modulator. Although the apparent advantages of SSB over DSBTC-AM are obvious, the circuitry required to demodulate the SSB waveform is considerably more complex than DSBTC-AM demodulators. The transmitter circuitry needs to be linear for SSB transmission while DSBTC-AM transmitters can be non-linear if the modulation is performed at the final or high power stages. The additional requirements in both the receiver

and transmitter sections in a SSB system more than double the unit costs as compared to DSBTC-AM systems.

Departing from conventional amplitude modulation another common modulation format is frequency modulation (FM). FM can have the form of narrowband or wideband FM. The difference lies in the magnitude of the modulation index. The magnitudes of the spectral components of the frequency modulated carrier can be seen by expanding the basic time domain representation of the FM wave into a sum of Bessel functions of the first kind. The fundamental function can be represented by the relationship,

$$f_c(t) = \cos(\omega_c t + \beta \sin \omega_m t)$$

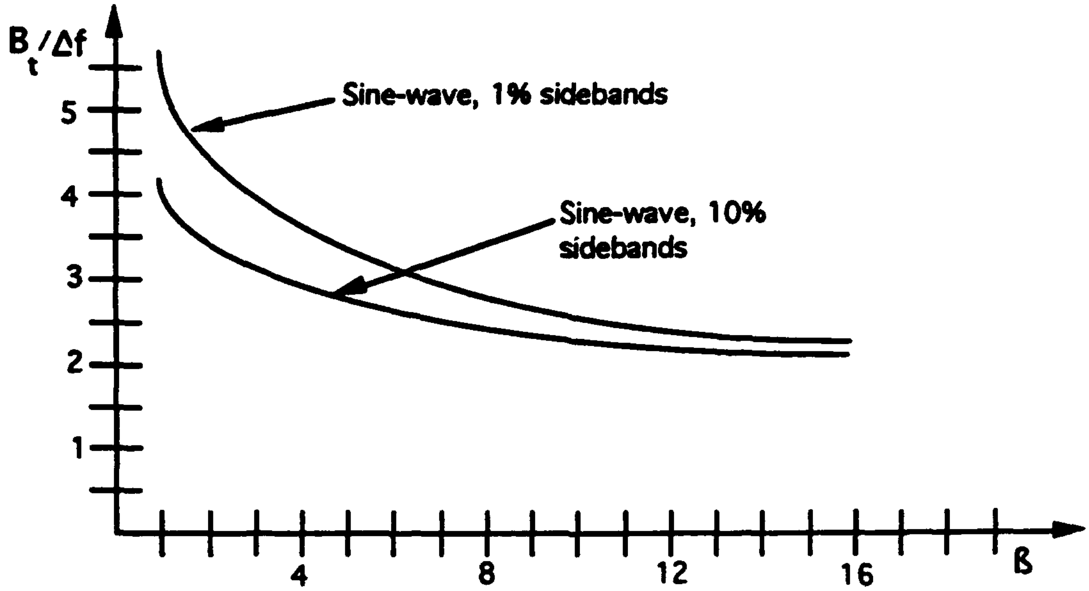
with

$$\beta \equiv \frac{\Delta\omega}{\omega_m} = \frac{\Delta f}{f_m}$$

or expressed in terms of Bessel functions of the first kind,

$$\begin{aligned} f_c(t) = J_0(\beta)\cos(\omega_c t) &- J_1(\beta)[\cos(\omega_c - \omega_m)t - \cos(\omega_c + \omega_m)t] \\ &+ J_2(\beta)[\cos(\omega_c - 2\omega_m)t + \cos(\omega_c + 2\omega_m)] \\ &- J_3(\beta)[\cos(\omega_c - 3\omega_m)t - \cos(\omega_c + 3\omega_m)] \\ &+ \ldots \end{aligned}$$

For small values of β ($\beta << \pi/2$), the only Bessel functions of significant magnitude are $J_0(\beta)$ and $J_1(\beta)$. This is the case for narrowband (NB) FM where in the frequency domain the bandwidth is approximately twice the highest baseband frequency. As β increases the magnitudes of the higher order terms contribute while the magnitude of the carrier begins to decrease. Spectral occupancy, or transmission bandwidth, B_t, normalized to the baseband bandwidth, Δf, of a FM signal as a function of β, is graphically illustrated below.



The transmission bandwidth is often approximated by the "rule-of-thumb" relationship below.

$$B_T = 2\Delta f + 2B = 2B(1 + \beta)$$

Alternate forms of analog modulation types include single sideband (SSB), vestigial sideband (VSB), and quadrature amplitude modulation (QAM). These techniques offer at least twice the channel capacity available with DSBTC-AM or NBFM. However, each involves a significant increase in circuit complexity over the opportunities available with digital modulation formats. Table 6-1 in the previous section summarizes the semi-quantitative performance tradeoffs between each type of analog modulation.

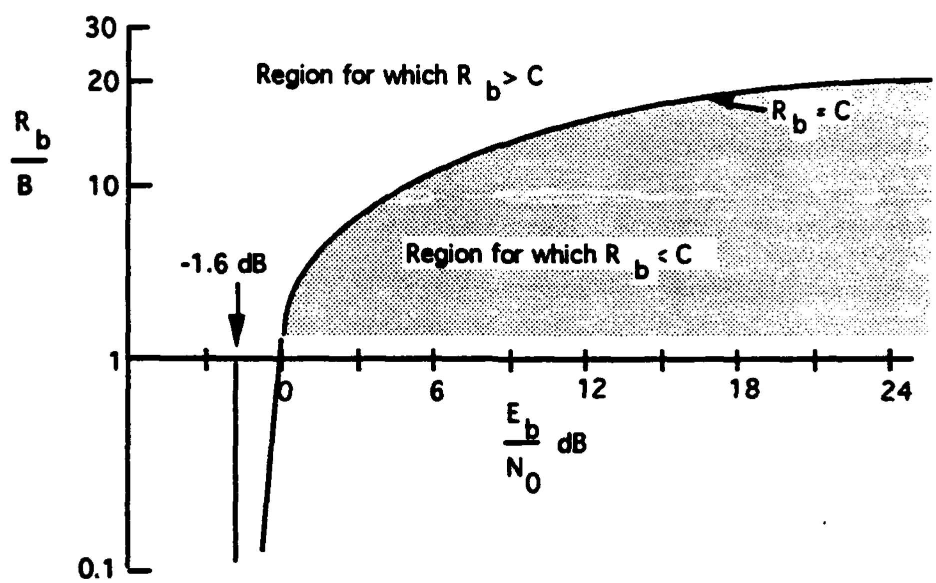
Digital modulation techniques are an attractive alternative to the various analog methods. The combination of increased channel capacity and the possibilities of multiplexing offer significantly higher information rates over analog methods. The ideal digital system transmits data at a bit rate R_b equal to the channel capacity C. If E_b is the received (or transmitted) energy per bit, the energy-per-bit to noise power spectral density ratio in terms of the bandwidth efficiency for an ideal system can be defined by the equation below.

$$\frac{E_b}{N_0} = \frac{2^{C/B}-1}{C/B}$$

Recall that the channel capacity, C, of a frequency bandwidth B (Hz), perturbed by additive white Gaussian noise of power density $N_0/2$ is given by [1]

$$C = B \log_2 \left(1 + \frac{P}{N_0 B}\right) \text{ bits/s}$$

where P is the average received (or transmitted) power. The above relationships set the stage for relative comparisons of various digital modulation methods. A plot of R_b/B versus E_b/N_0 highlights the tradeoffs between the various system parameters limited by the physical design. This plot is typically referred to as the bandwidth efficiency diagram shown below.

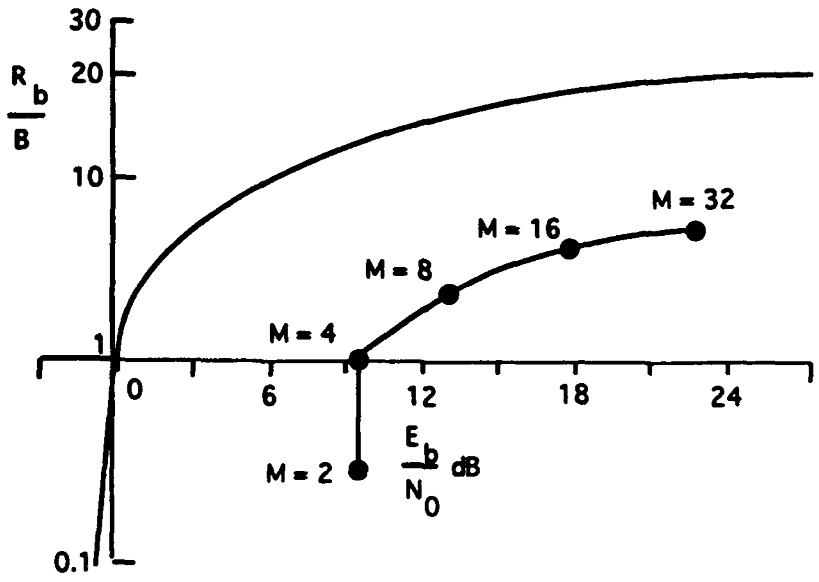


Assigning the definitions of bandwidth efficiency for M-ary phase shift keying (PSK) and M-ary frequency shift keying as [2]

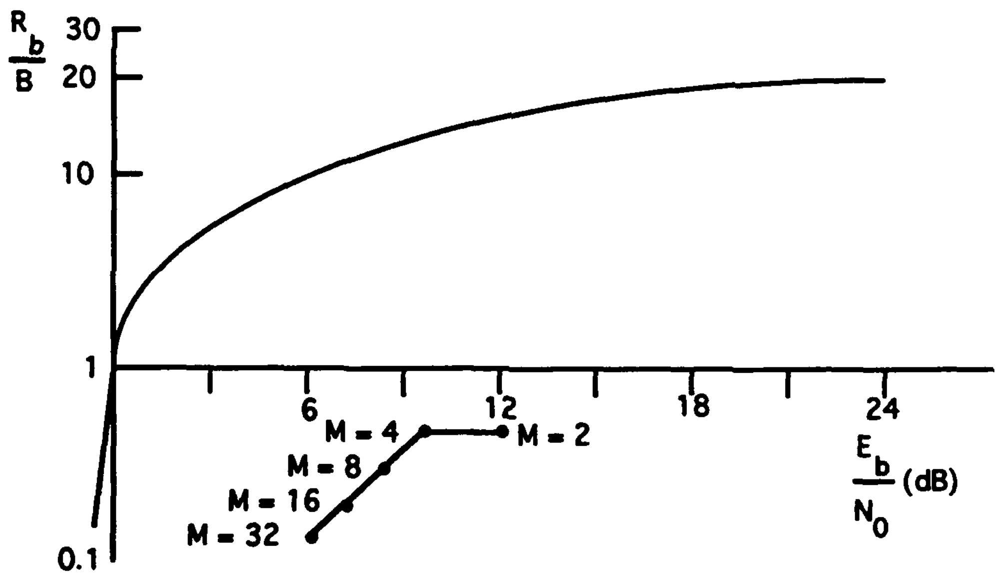
$$\frac{R_b}{B} = \frac{\log_2 M}{2} \qquad \text{for M-ary PSK}$$

$$\frac{R_b}{B} = \frac{2 \log_2 M}{M} \qquad \text{for M-ary FSK}$$

The performance of each can be compared against the ideal digital system as shown in the bandwidth efficiency diagram below.



M - ary PSK



M - ary FSK

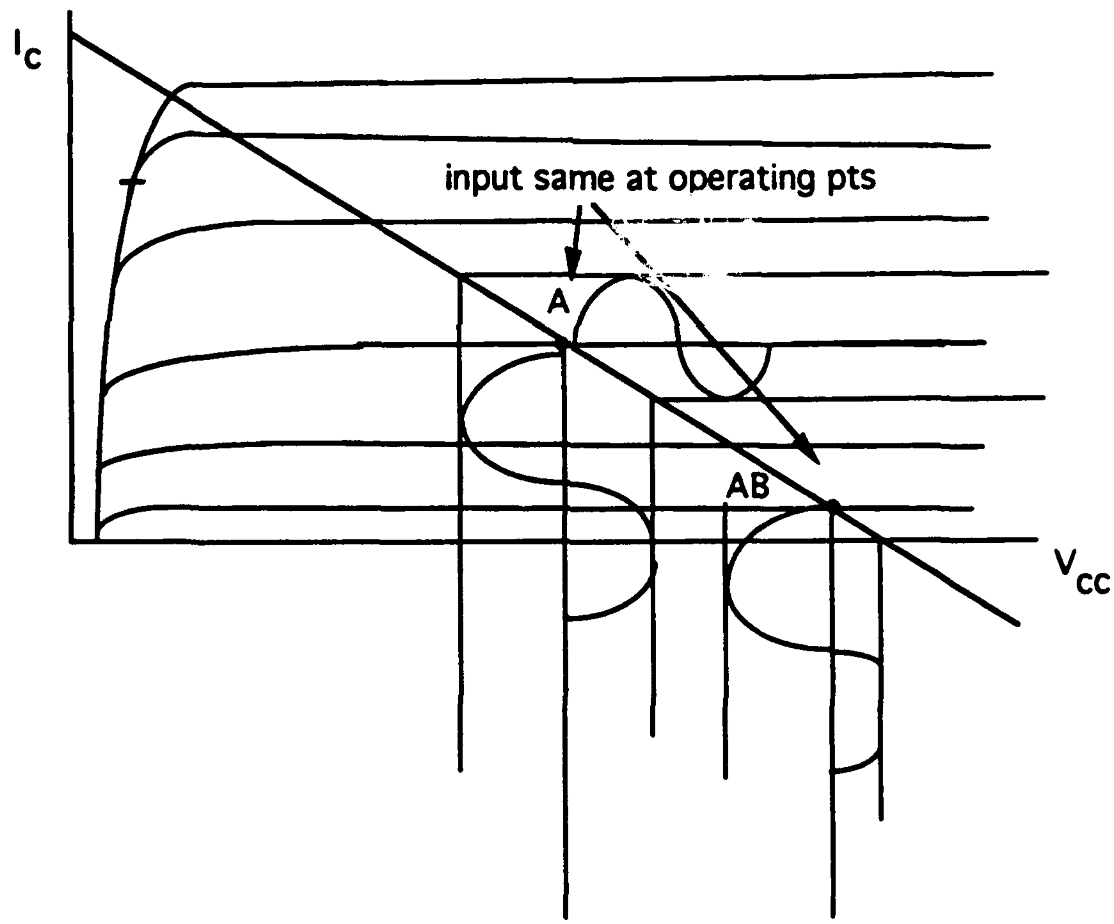
$P_e = 10^{-5}$ for both charts

The type of modulation selected depends on the system requirements and application of interest. When simplicity and economy are more important than bandwidth efficiency, binary FSK with noncoherent detection is a common choice. The maximum bit rate would be approximately half the bandwidth while the frequency shift would be one-half to three quarters the maximum bit rate. Noting the spectral inefficiencies with FSK formats, other modulation formats are chosen if high information rates are needed and the bandwidth is limited. The higher efficiencies offered by M-ary PSK or M-ary QAM formats are commonly utilized in digital radio systems where rigid transmission bandwidths are specified. In M-ary PSK systems the in-phase and quadrature components are structured to maintain a constant waveform envelope, while in QAM systems the carrier experiences amplitude as well as phase modulation. PSK systems have the drawback of being susceptible to phase errors or jitter, decreasing the isolation between each phase or channel state. Higher phase order PSK systems have limited bit rates primarily due to the degree of phase error present on the signal prior to demodulation. QAM formats offer a partial solution to the phase jitter problems (and resulting information rate limitation) associated with PSK. Each signal point is isolated both in phase and amplitude. The transceiver however is burdened with more complex and expensive clock and signal recovery circuits. The nonconstant envelope places additional linearity requirements on both the RF transmitter circuitry and the RF and IF down conversion components. The radio link is also more susceptible to channel imperfections such as multipath and fading.

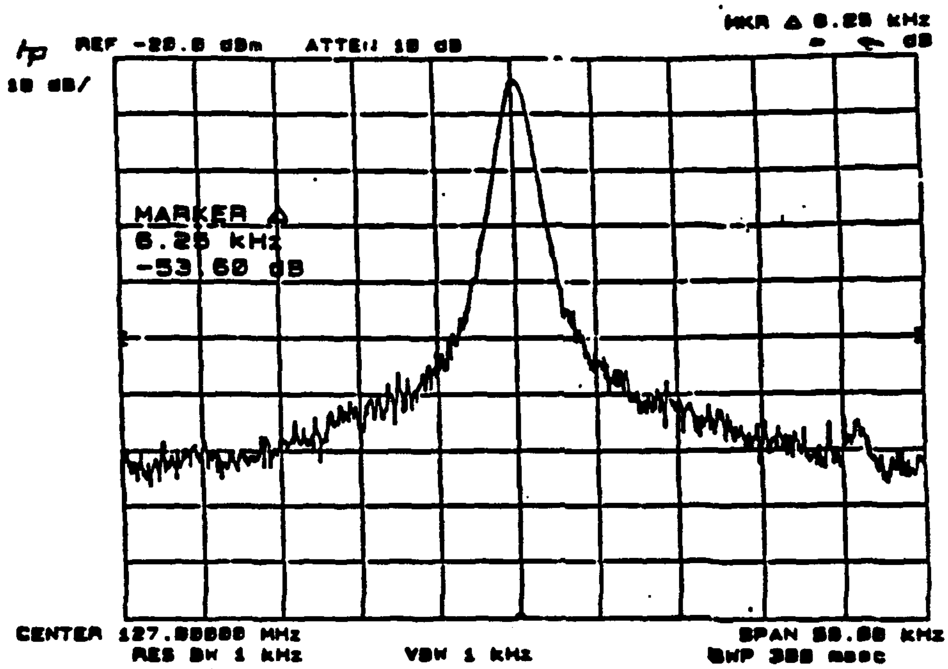
Transmitter design

The degree of extraneous sideband energy related to the modulation, the level of non-thermal noise about the carrier, and any additional spurious products due to the transmitter will be a function of the operating point of the amplifier. Modulation techniques such as AM and SSB require a linear amplifier, but are insensitive to phase characteristics of the amplifier. Frequency modulated waveforms however, do not require linearity, but can be sensitive to delay or phase variations across the information bandwidth. Therefore the size, power consumption, and eventual cost of the the transmitter sections are directly related to the modulation technique implemented. Allocating closer channels, in itself, has no significant impact on the transmitter design directly. However, if this change is coupled with a redesign of the current system's architecture, then the transmitter design is affected.

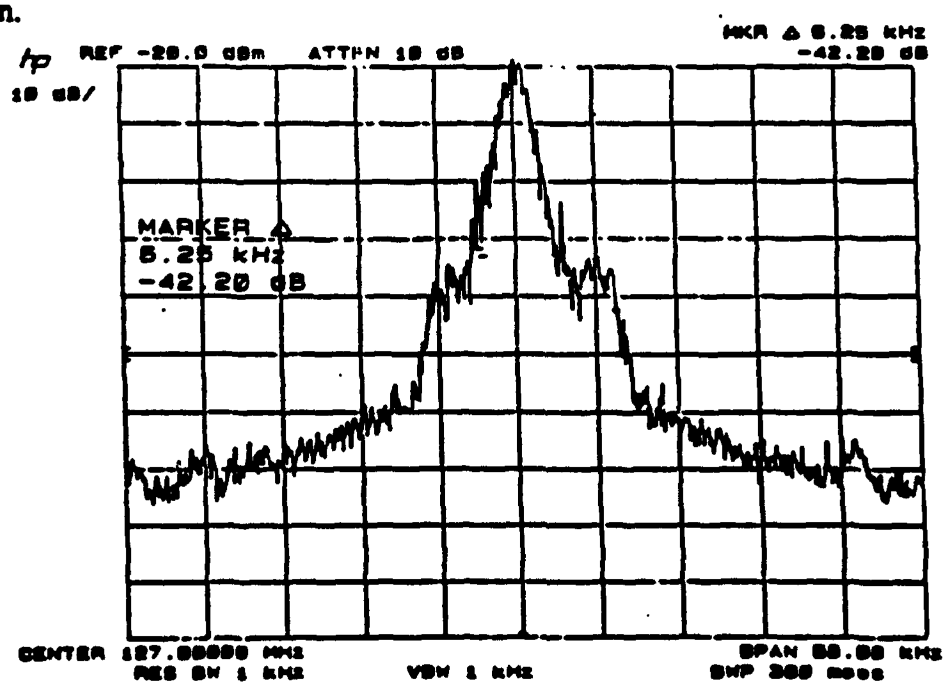
The amplifiers found in most airborne and ground AM transceivers will of class AB or C, with the later found in the airborne systems. Recalling that the efficiency of an amplifier is related to the current consumption over the period of the input RF waveform, the operating point determines the degree of linearity and the resulting efficiency. Ideal class A designs are biased over the complete period of the applied signal and offer the necessary dynamic range to reproduce the sinusoidal waveform, but at the expense of low efficiency. Harmonics and unwanted spectral products from an amplifier occur when a design departs from the ideal case and the input signal is not completely reproduced at the output. This is more clearly illustrated by the figure below, which plots nominal operating current versus voltage (I vs V) curves for conventional output power devices, highlighting the various operating points.



The distortion products are evident in most AM transceiver designs as a broad pedestal about the carrier. Below are spectral plots of a VHF transmitter designed for portable military ATC. The first plot is the spectrum of an unmodulated carrier.



The second spectral plot is the same carrier modulated with voice. Note the +/- 5 kHz side lobes at -35 dBc. The modulation bandwidth is limited to a 4 kHz bandwidth centered about the carrier. The secondary sidelobes and broadband noise evident in the spectra are an example of the potential sources of interference for a receiver configured for narrowband operation.



Frequency Stability

The dominant cost driver in the frequency budget is the synthesizer stability. As discussed previously most commercial avionics utilize crystal controlled oscillators with stabilities of 10 to 30 ppm which translates to a maximum frequency offset of approximately 4.1 kHz at 137 MHz. The total frequency bandwidth needed could be as high as 10.24 kHz[2]. An improved reference oscillator would reduce this allocation significantly. The next level in oscillator performance is found with Temperature Compensated Crystal Oscillators (TCXOs). The stabilities of 1 to 5 ppm offered by TCXOs would reduce the needed bandwidth by half, but would increase the system cost associated with the oscillator proportionally. The ideal system would be to have a synthesizer stability of +/- 1ppm and an high order IF filter with 3 dB bandwidth of 5 kHz. This ideal system is not practical, since the ability to detect present day systems would be completely lost. Clearly the first order tradeoffs are final system cost and backward compatibility.

The tradeoffs associated with meeting the required oscillator stability is related to the specific market a new radio will serve. The cost of a TCXO is not prohibitive in relation to the commercial aviation (CA) industry. The CA user has increased visibility to the evolving problems and issues associated with modern day ATC. As traffic densities slowly increase the work loads and frequency assignment issues become relatively more apparent. While present CA radios are in the $9000 to $15,000 range, their GA counterparts are in the $1200 to $3000 range.

One method to alleviate the financial burden of TCXOs in the airborne transceivers is a beacon or frequency standard transmission system. A network of ground transmitters relay coded signals based on a highly accurate frequency reference oscillator. The airborne transceivers would process the received frequency reference and re-align their internal oscillators. Since the spatial coverage of a properly configured ground network would encompass many airborne users, it is possible that the ground installation costs could be significantly overshadowed by the savings of a lower cost oscillator aboard potentially thousands of aircraft. For example, a large commercial airline with an field inventory of 500

---

2 Total bandwidth needed is the sum of the information bandwidth, 6 kHz, the maximum Doppler offset, 0.14 kHz, and the overall stability factor, 4.1 kHz.

transceivers would save over $50,000 in direct costs through distributed frequency sytsem. The costs associated with a ground based distribution system are in part based on tax dollars, with the remainder spread over multiple users.

The actual format of a distributed reference oscillator or "timing system" can take many forms. The most widely used timing formats include:

1. *Side-tone ranging (STR) signaling* A sinusoidal carrier is amplitude or phase modulated by a sequential or simultaneous set of tones. A composite set of tones would have the form,

$$m(t) = \sum_{i=0}^{N} a_i \sin(2\pi f_i t)$$

The frequency tones are of the form,

$$f_i = f_0 4^i$$

The period of each tone is measured and compared against the fundamental.

2. *Pseudonoise ranging signal* A sinusoidal PSK or QPSK signal is modulated by one or more pseudonoise sequences, whose period equals or exceeds a predetermined value T.

3. *Time-gated pseudonoise ranging signal* A signal of the same form as the preceding signal except it is time gated allowing for short duty factor.

Each of the above formats is based on the transmission of a known signal, $s(t + \tau_0)$, and the reception of, $s(t + \tau + \tau_0)$, where τ is the propagation delay.

7.5 IMPACT/IMPORTANCE

It would seem that the principal benefit of proposing closer channel spacing as a solution to a capacity problem for ATC communications would be the ability to be compatible with previous generations of communications equipments. The circuitry needed

for dual detection does not possess any technological challenges even though there is a significant cost impact on the low-end GA class of radios. The existing GA class VHF radios could not be "upgraded" to support 12.5 kHz operation by simply changing the reference oscillator and last IF filter. All of the parameters discussed above: loop bandwidth, VCO phase noise, reference oscillator stability, etc., need to be significantly improved to support 12.5 kHz operation *and* maintain full use of the additional channels. The next generation VHF ATC system will possibly utilize advanced modulation and coding techniques to maximize the spectral efficiency beyond simple channel splitting. This would impose additional requirements on additional system modules such as the transmitter, power supply, and system processor.

The cost of the 1 to 5 ppm oscillator stability requirement needed to support a next generation system is currently at $100 in large volumes. As an alternative, one could use higher grade oscillators in the ground controller radios and a time distribution system. Airborne transceivers could realign their reference oscillators based on the accurate time distributed by the ground network. The burden placed on the airborne transceiver would be to provide the necessary processing to allow for clock alignment. A radio based on a digital waveform would have the microprocessor needed but an analog system typically does not.

The second significant cost impact on an upgraded transceiver is the spectral characteristics of the transmitted waveform. Since the nearest possible channel would be 12.5 kHz away , the spectral spillover of an adjacent channel must be held at a level to minimize interference. The level of the close-in spurious responses of the synthesizer and the transmitter must be below the dynamic range of the front-end componentry of a nearby victim receiver. The costs associated with designing a transmitter which could meet these requirements are constrained to the initial design phases of the development , but not directly related to the final unit production costs.

The most significant impact of closer channel spacing does not involve the radio designer or the end user. The evolution from the current communication system to a more modern one involves significant frequency management changes and the possibility of revamping the ground radio sites. The current radios are configured with separate units for the transmitters and the receivers. The corresponding antenna elements for each unit are configured with separate antenna towers for primary transmitters and primary receivers.

Typical antenna separations are specified to be no less than 80 ft. A system which involves integral transmitters and receivers or transceivers, would not be able to take advantage of the physical isolation provided by the antenna separation unless the transmissions are carefully orchestrated with the site configuration. Frequency allocations for each site should be examined and coordinated with the actual location of each transceiver. Strong signals in close proximity to one another can cause significant degradation of many channels beyound the ones in use. This is primarily due to receiver desense and IMPs[3]. In the same manner that improperly designed internal circuitry can cause self interference, improperly configured transceivers at a common location can cause similar interference problems.

7.6 TRANSITION

The next generation ATC system will offer the users increased capabilities which are currently under examination and are addressed by the various branches in the decision tree. Each parameter or technological feature of a new ATC system which aids in providing this increased performance adds to the degree of circuit complexity. Circuitry optimized for advanced modulation and coding techniques coupled with basic modern day radio design as compared to the relative simple circuitry needed for conventional AM, warrants a dual mode radio to accommodate the present day waveform and associated users. The circuitry needed to support transmission or reception of DSBTC-AM could be placed on separate PCBs, which later could be removed when no longer needed. Two options exist in the design which would account for a gradual introduction of a closer channelized system while allowing for existence of the present day radios.

The first option would be to design a 12.5 kHz channelized system based on the chosen modulation and user requirements. Specific components such as the IF filters, reference oscillators, etc., would be specified for optimal performance along a main path in the receive chain. Backward compatibility would be accomplished via a secondary path in the receive chain branched at the last IF. This path could be selected after it was determined that conventional DSBTC-AM was present in the main channel (or adjacent channels). Modulation recognition is typically performed using digital signal processing techniques. Algorithms perform statistical tests on the sampled waveform using both the envelope and

[3] Refer to appendix B for a detailed discussion of the ground radio co-location issues.

zero-crossing characteristics of the signal. Most algorithms optimized for modulation identification are based on the *concept -of-similarity* developed in pattern recognition [3]. The algorithms are simple and readily provide high confidence for moderate signal-to-noise ratios.

The most cost effective method to accomplish a dual detection capability would be to provide a separate AM detection module at the last IF. The combined integrated circuitry for detection and a monolithic crystal filter could be integrated on printed board area of less than two square inches. Upon complete transition to the new system architecture the module could be removed or no longer supplied in the case of a new purchase.

The second method would be to specify the new system components to account for the tolerances of the older systems. Although the new receiver could possess increased stability and finer frequency resolution, the pre-detection bandwidth would need to be approximately 10 to 12 kHz wide. This naturally would not allow for the 70 dB adjacent channel rejection desired by certain members of the ATC communications community. The increase in available channels offered by a closer channel spacing could be preserved in low to moderate traffic densities by prudent frequency management. Both frequency re-use and adjacent channel use could be effectively orchestrated given the spatial arrangements of the frequency assignments. This method provides a limited degree of increased performance at the benefit of lower cost.

The transmitter could be designed such that the operating point of the intermediate and final amplification stages could be selectable as a function of the modulation being used. This technique is very common in military communication systems where multi-mode waveforms are a requirement and space is at a premium. Since relatively few extra devices are needed to support a dual modulation capability in the transmitter, the cost and reliability impact is negligible.

The existence of older radios, which were designed to the tolerances associated with 25 kHz or 50 kHz channelization, must be considered in the frequency management during a transitional period. The broader signal spectra and frequency drifts of the older AM radios will cover several 12.5 kHz channels. Although the new radio could be designed to detect the AM waveform, the additional channels gained with a 12.5 kHz allocation would be lost in

the frequency spectrum where AM is still allowed. Separate frequency allocations within the 19 MHz for AM and the "new" waveform would aid in isolating the effects of interference. The number of additional channels gained through a narrower allocation would naturally be limited to the frequency bands chosen for the new waveform. The users would slowly transition to a new transceiver as avionics or platforms are replaced or upgraded. The AM frequency allocation would slowly decrease as the number of users decreases.

7.7 CRITERIA FOR DECISION

The criteria implied by the above discussions falls into three major categories: production item unit cost, user requirements, and the desire for a smooth transition from the systems of today.to the ones of the future. The ability to gain broad user acceptance of a next generation communication system is directly related to providing an enhanced capability which satisfies an existing deficiency without burdening the end user financially. In the GA environment the user requirements and unit costs are proportionally lower to their counterparts in CA.

The ability to provide closer channels within the ATC communications bands is related to the quality and characteristics of specific receiver components such as IF filters, synthesizers, etc., and the future anticipated volume of controlled air space. The latter sets the stage for the future system. The anticipated service volumes, projected over the average system lifetime, will determine the degree of spectral efficiency needed versus the unit cost. Resolving the future needed capacities into firm requirements allows for definition of the many system parameters such as the modulation type, which in turn determines the specific component characteristics. The transition and/or integration phase of the new system will directly determine the use these components and the degree of circuit complexity.

The criteria for decision can be grouped into two segments: the requirements of the GA user, and the requirements of the CA users. Both set of requirements will in turn determine the end unit cost. Spectral congestion and the desire for more information channels is more visible in the European commercial theater but can easily extrapolated to the conditions of future CA communications in the United States. The inntegration and use of a communications system also determine the specific unit characteristics. Parameters such as

adjacent channel rejection, selectivity, sensitivity, etc., in a receiver are directly related to the channel bandwidth but are also related to the number of users within a coverage volume.

7.8 CONNECTIVITY/RELATIONSHIP WITH OTHER DECISIONS

Although this paper provides insight to the factors involved in the design of a transceiver with closer channel spacing compared to its predecessor, the outcome of several branchpoints in the tree resolve many of the above issues. The complexity of the chosen modulation not only determines device characteristics in the receive chain which relate to selectivity and increased channel capacities but it also determines the level of processing power needed. The choice of narrow channel bandwidth eliminates certain "leaves" and branches in the tree since certain options are incompatible with narrow bandwidths, such as wideband FM and various high data rate modulation formats.

Within the branch of spectrum utilization each sub-branch is directly related to the channel bandwidth and the radio design tradeoffs associated with this bandwidth. Communication throughput is directly affected by the chosen bandwidth, since the performance offered by advanced modulation techniques is directly related to the allocated channel bandwidth. As demonstrated in the above discussions, the channel spacing is a significant variable in the overall system performance.

SECTION 8

DIGITAL MODULATIONS

8.1 CONTEXT

A systematic process for examining technical alternatives for improved air/ground communications in air traffic management has been established. A decision tree structure is shown in Appendix A that attempts to organize various alternatives in a top-down hierarchy. This provides a framework for evaluating potential solutions that can be represented by paths through the decision tree.

This paper addresses a particular alternative in the decision tree, namely, branchpoint 4.1.2.1, (Digital) Modulation Scheme. The organization of this paper is as follows: Background, Issues, Tradeoffs, Impact/Importance, Transition, Criteria for Decision, and Connectivity/Relationship with Other Decisions. These alternatives are not developed in a vacuum but with sensitivity to the realities of air traffic management. Cost is treated in this paper only in terms of relative implementation complexity because of limited resources.

8.2 BACKGROUND

Current VHF air/ground (A/G) communications in the 118-137 MHz frequency band use the Double Side Band Transmitted Carrier (DSBTC) form of analog modulation (see section 6). The capabilities of the current system for digital data transmission are limited. A Minimum Shift Keying (MSK) digital baseband waveform operating at only 2400 b/s is used to modulate an AM radio in the Aircraft Communication Addressing and Reporting System (ACARS). Also, there are no automatic frequency changes by in-band signaling or other means, nor semi-automatic handoff of aircraft from one controller to another; these feature capabilities are most readily achieved with digital signaling waveforms. The far-term improvement of VHF A/G communications clearly lies with digital modulation. Although both advanced analog, e.g., single sideband, modulation and digital modulation can provide higher throughput, digital modulation and coding techniques can achieve much more reliable transmission, alternate routing/networking and other features that are difficult if not impossible to implement in analog systems. Digital modulation techniques accept a sequence of numbers and then translate them into discrete amplitudes, offset phases,

frequency deviations, or combination of these in modulating a transmitted radio frequency carrier. Data is represented by a sequence of numbers and in the case of voice, the speech waveform is transformed into a sequence of numbers by a digital vocoder.

The motivation for considering digital modulations for VHF A/G communications arises from the following objectives:

1. To improve channel bandwidth efficiency, i.e., more information received per unit bandwidth.

2. To replace and upgrade the current analog radios, which may become logistically unsupportable in the future.

3. To combine voice and data capabilities in a single radio.

4. To improve power efficiency, i.e., lower signal-to-noise ratio required to achieve a reliable communication link.

5. To increase the resistance to external sources of interference, such as man-made noise, cosited radio interference, and co-channel and adjacent channel interference. Adjacent channel interference is a major impediment in reducing the 25 kHz frequency spacing in attempting to provide more channels with the existing analog modulation. (This topic is treated in section 7.)

6. To reduce susceptibility to multipath fading, especially at longer ranges and with rotary wing aircraft.

7. To provide data link capabilities. The current ACARS data rate needs to be increased so that transmitted packets are shorter and less likely to interfere with other packets in the same channel shared among many aircraft.

8. To permit interoperability with other communication entities and electronic systems through the Aeronautical Telecommunications Network (ATN).

Digital modulation can address all of these needs, especially items 2, 6, 7 and 8, much more readily than analog modulations.

This decision tree paper considers various candidate digital modulation techniques for replacing the current analog modulation in the future A/G system for Air Traffic Services (ATS) and Aeronautical Operational Control (AOC). First, each of the important issues used in comparing different modulation techniques is discussed. Then the pros and cons of various advanced digital modulation techniques are presented.

8.3 ISSUES

Major factors to be taken into account in comparing different digital modulation techniques are related to improved A/G communication capabilities which can be defined by specific technical and operational criteria. System tradeoffs are fundamental to all digital communication designs. The issues involved are:

1. Bit-error-rate (BER) performance defined as the information signal-to-noise ratio, E_b/N_o, required to achieve a certain information bit-error-probability, P_{be}. This is a measure of a digital modulation scheme's power efficiency.

2. Spectral compactness of the modulated signal. The shape of the power spectral density suggest a signal's potential interference effects on other signals and with itself. This is a measure of the susceptibility to interference, such as intersymbol interference (ISI), co-channel interference (CCI) and adjacent channel interference (ACI). A minimum channel bandwidth, W, for a specific signal-to-interference ratio can be determined.

3. Bandwidth efficiency, the information transmission rate in bits per second per Hertz (b/s/Hz) of a suitably defined signal bandwidth.

4. Effect of nonlinearities, a measure of system degradation from the theoretical ideal due to implementation with the inevitable use of nonlinear electronic devices in the receive and transmit elements of the system.

5. System robustness measured by the performance degradation caused by both slow and fast fading in the A/G communications channel.

6. Implementation complexity required to transmit and receive the digital modulation.

Channel bandwidth and transmitted power constitute two primary digital communication resources. Both are closely related to the issues listed above. The primary objective of spectrally efficient modulation techniques is to maximize bandwidth efficiency.

Bandwidth efficient modulation techniques include phase shift keying (PSK), quadrature amplitude modulation (QAM) and continuous phase frequency shift keying (CPFSK). The basic digital modulation techniques include simple binary shift keying (branchpoint 4.1.2.1.1), quaternary shift keying (branchpoint 4.1.2.1.2) and M>4-ary shift keying (branchpoint 4.1.2.1.3). PSK signals include binary PSK (BPSK), quaternary PSK (QPSK) and its derivatives, 8-ary PSK (8-PSK) and 16-ary PSK (16-PSK). The quadrature scheme of QPSK is explained in more detail because of its practical importance. Its modification to other modulations, such as Offset QPSK (OQPSK), $\pi/4$ shifted QPSK ($\pi/4$-QPSK), Aviation QPSK (A-QPSK), Minimum Shift Keying (MSK), and Gaussian filtered MSK (GMSK) are discussed. Two classes of QAM signals, 4-ary Offset QAM (4-OQAM) and 16-ary Offset QAM (16-OQAM) are included. Finally, for FSK signals, we include binary FSK (BFSK) and 8-ary FSK (8-FSK) in the discussion below. The channel signal constellations for these modulation techniques and their corresponding Gray codes are shown in figure 8-1.

Binary Shift Keyed Signaling

In binary shift keying, only two signals are used in the original data stream. BFSK and BPSK are two binary modulations now widely used in communications systems.

In BPSK modulation, the modulating data signal shifts the phase of a carrier to one of two states, either 0 or π. In a polar plot of signal representation in which the vector length corresponds to the signal amplitude and the vector direction corresponds to the signal phase, the two possible BPSK signals can be depicted as opposing vectors and thus are called

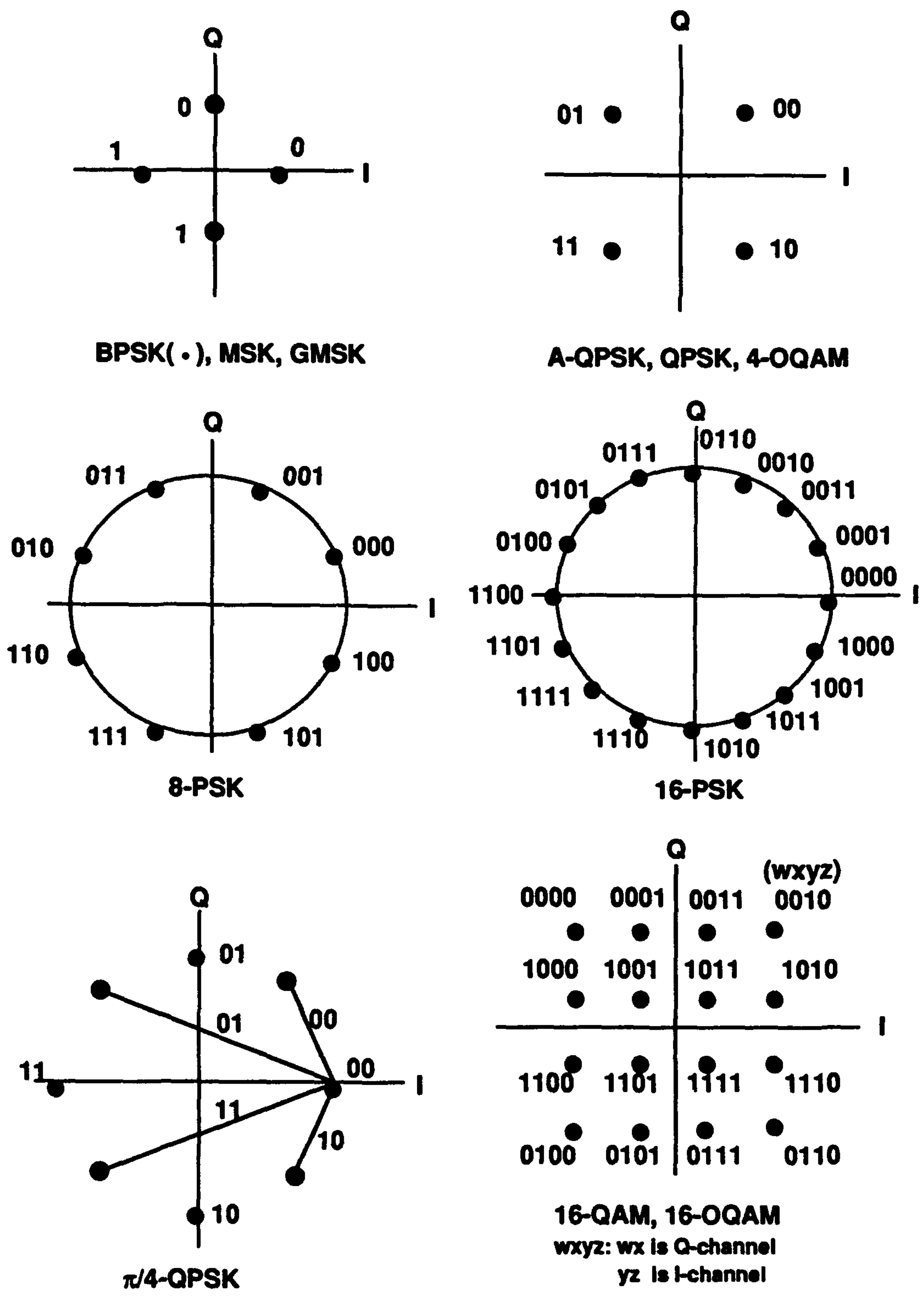


Figure 8-1. Channel Signal Constellations and Gray Codes

antipodal signals. When BPSK is detected coherently a replica of the transmitted carrier with proper frequency and phase is needed at the receiver. Since noncoherent means detection without carrier phase information, the noncoherent receiver of BPSK performs a phase comparison detection operation by deciding whether the currently received signal phase has changed compared to the preceding demodulated phase. This is also called differential phase shift keying (DPSK). DPSK eliminates the need for a coherent reference signal at the receiver by combining two basic operations at the transmitter: 1) differential encoding of the input binary data; and 2) phase shift keying. The receiver is equipped with a memory capability, so that it can measure the relative phase difference between the waveform received during two successive bit intervals. DPSK can also be considered as noncoherent orthogonal modulation over two bit intervals.

In BFSK modulation, the modulating data stream can shift the frequency of the carrier between a positive and a negative offset. This results in an abrupt frequency change at the bit transition. In order to preserve the orthogonality of the two possible frequencies (tones) during the detection, they must have a frequency separation of a multiple of 1/2T or 1/T Hertz [1], where T is the bit duration. The minimum required spacing for noncoherently detected BFSK signaling is 1/T; while the minimum spacing for coherently detected BFSK signaling is 1/2T. Therefore, for the same symbol rate, coherently detected BFSK can occupy less bandwidth than noncoherently detected BFSK and still retain orthogonal signaling. Thus coherent BFSK can be more bandwidth efficient. However, in practice, BFSK transmitters typically do not control the initial phase of a tone, and the necessary phase tracking at the receiver can take several bit intervals, so the 1/2T spacing is rather impractical to achieve.

M-ary Shift Keyed Signaling

In M-ary signaling, each elementary signal carries $k = \log_2 M$ bits of information for alphabet size M = 2, 4, 8, 16, . . ., etc. The processor handles k bits at a time from the input data stream and instructs the modulator to produce the corresponding waveform from the set of M waveforms. In M-ary PSK, the M waveforms are represented by a set of uniformly spaced phase angles; while in M-ary FSK, the M symbols usually correspond to M tones equally spaced in frequency by 1/2T or 1/T depending on whether coherent or noncoherent

detection is used. Similarly to DPSK, as described in the last section, M-PSK can be detected noncoherently. Its noncoherent version, phase comparison (differential) M-PSK, can be obtained by differentially encoding the M-ary symbol waveforms in the transmitter.

As an example of applications, 8-PSK has been widely used in the 6 GHz and 11 GHz high capacity (90 Mb/s) microwave digital radio system [2, 3].

QPSK and OQPSK Signaling

Compared to BPSK, QPSK and OQPSK are two bandwidth conserving modulation schemes for the transmission of digital information. The original data stream $d(t) = d_0, d_1, d_2, d_3, \ldots$, consists of bipolar rectangular pulses of value +1 or -1, representing binary one ("1") and zero ("0"), say, and is divided into an in-phase stream $d_I(t)$, and a quadrature stream, $d_Q(t)$ with pulse duration doubled as follows:

$$d_I(t) = d_0, d_2, d_4, d_6, \ldots$$
$$d_Q(t) = d_1, d_3, d_5, d_7, \ldots$$

so that $d_I(t)$ and $d_Q(t)$ each have half the bit rate of $d(t)$.

A QPSK waveform, $s(t)$, is achieved by amplitude modulating the in-phase and the quadrature data streams onto the cosine and sine functions, say, of a carrier frequency, f_0, as follows:

$$\begin{aligned} s(t) &= \frac{1}{\sqrt{2}} d_I(t) \cos(2\pi f_0 t + \frac{\pi}{4}) + \frac{1}{\sqrt{2}} d_Q(t) \sin(2\pi f_0 t + \frac{\pi}{4}) \\ &= \cos[(2\pi f_0 t + \theta(t)] \end{aligned} \qquad (1)$$

The pulse stream $d_I(t)$ amplitude-modulates the cosine function with amplitude of +1 or -1. This is equivalent to shifting the phase of cosine function by 0 or π; consequently, this produces a BPSK waveform. Similarly, the pulse stream $d_Q(t)$ modulates the sine function, yielding another BPSK waveform orthogonal to the cosine function. The summation of these two orthogonal components of the carrier yields the QPSK waveform. The values of $\theta(t)$, 0, $+\pi/2$, $-\pi/2$, or π correspond to one of the four possible combinations of $d_I(t)$ and

$d_Q(t)$. Since the sine and cosine functions are orthogonal, the two BPSK signals of equation (1) can be detected separately. QPSK is more spectrally efficient than BPSK; i.e., it transmits 2 bits per symbol instead of one bit per symbol at same symbol rate and same bandwidth.

OQPSK signaling can also be represented by equation (1); the difference between QPSK and OQPSK is only in the alignment of the two baseband waveforms. In QPSK, the odd and even pulse streams are both transmitted at the rate of 1/2T b/s and are synchronously aligned. In OQPSK, there is the same data stream partition and orthogonal transmission; the difference is that the timing of one of the I and Q pulse streams is shifted such that the alignment of two streams is offset by T. The unfiltered frequency spectrum as well as bit-error-rate (BER) performance of QPSK and OQPSK are the same. However, with OQPSK the maximum phase change at the bit transitions is 90^o while with QPSK, the maximum phase change at bit transitions is 180^o. The advantage of OQPSK is thus less spectrum spreading than QPSK due to system nonlinearities.

The QPSK modulation scheme found its application in the Avantek [2] and Farinon Electric [4] 1.7 to 2.3 GHz low capacity (3.152 Mb/s) *digital microwave data systems.*

A-QPSK Signaling

A-QPSK is OQPSK with pulse-shaping filters. The rectangular pulses of each data stream are passed through a shaping filter with an amplitude characteristic satisfying the mask shown in figure 8-2. The phase of the pulse shaping filters is linear to less than a $\pm 3^o$ deviation as shown in figure 8-3. The spectral output of the I- and Q- pulse shaping filters have sidelobes at least 40 dB down from the frequency center. A-QPSK is a standard modulation in the aeronautical-satellite community, i.e., International Civil Aviation Organization (ICAO), for channel data rates of 2.4 kb/s and above, and is used with the INMARSAT aeronautical-mobile system. One potential advantage of this scheme for VHF A/G communication is in the greater likelihood of ICAO acceptance. This would also help nurture the ultimate goal of a single avionics box for both A/G and satellite communications.

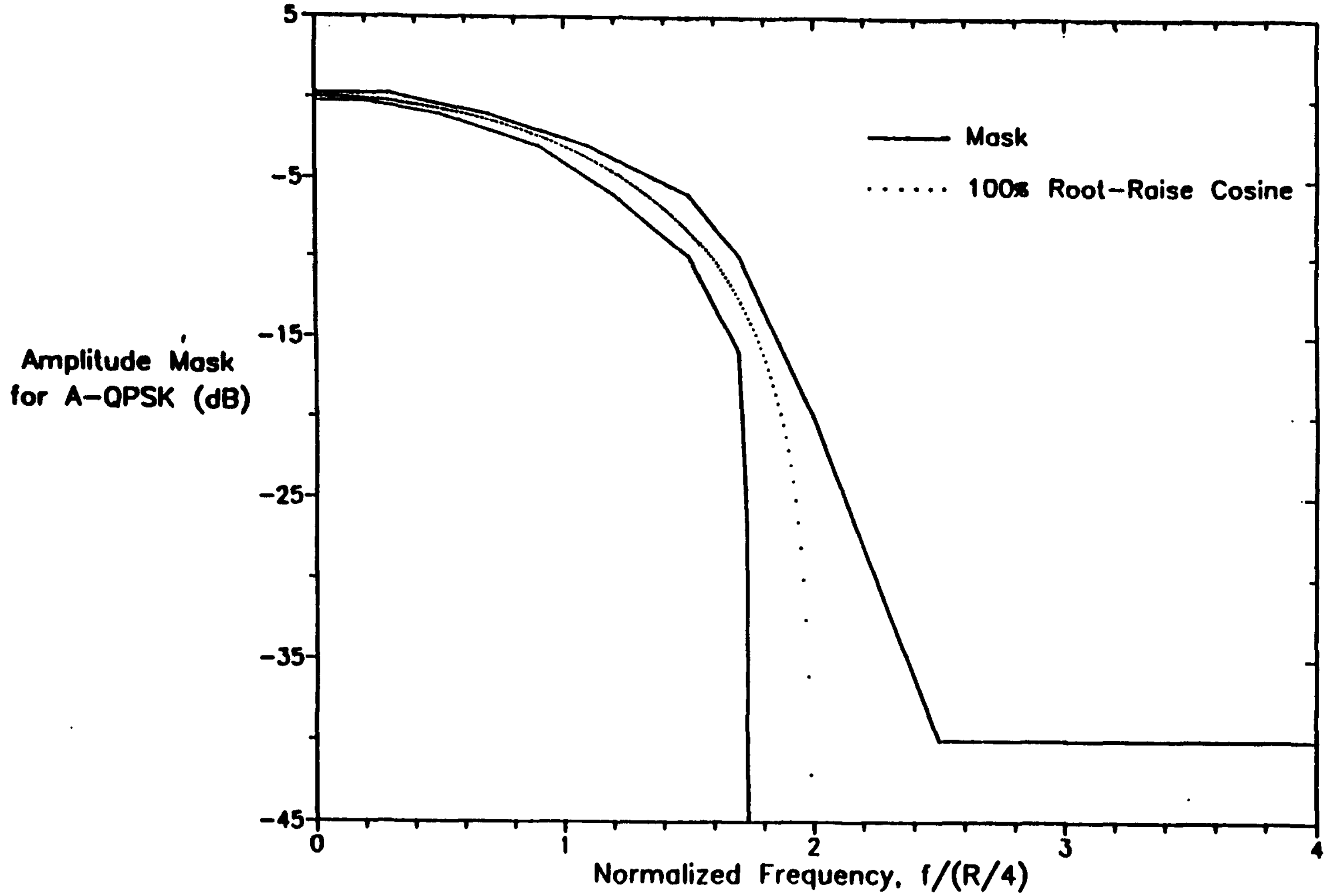


Figure 8-2. Amplitude Mask Requirement for A-QPSK Signals

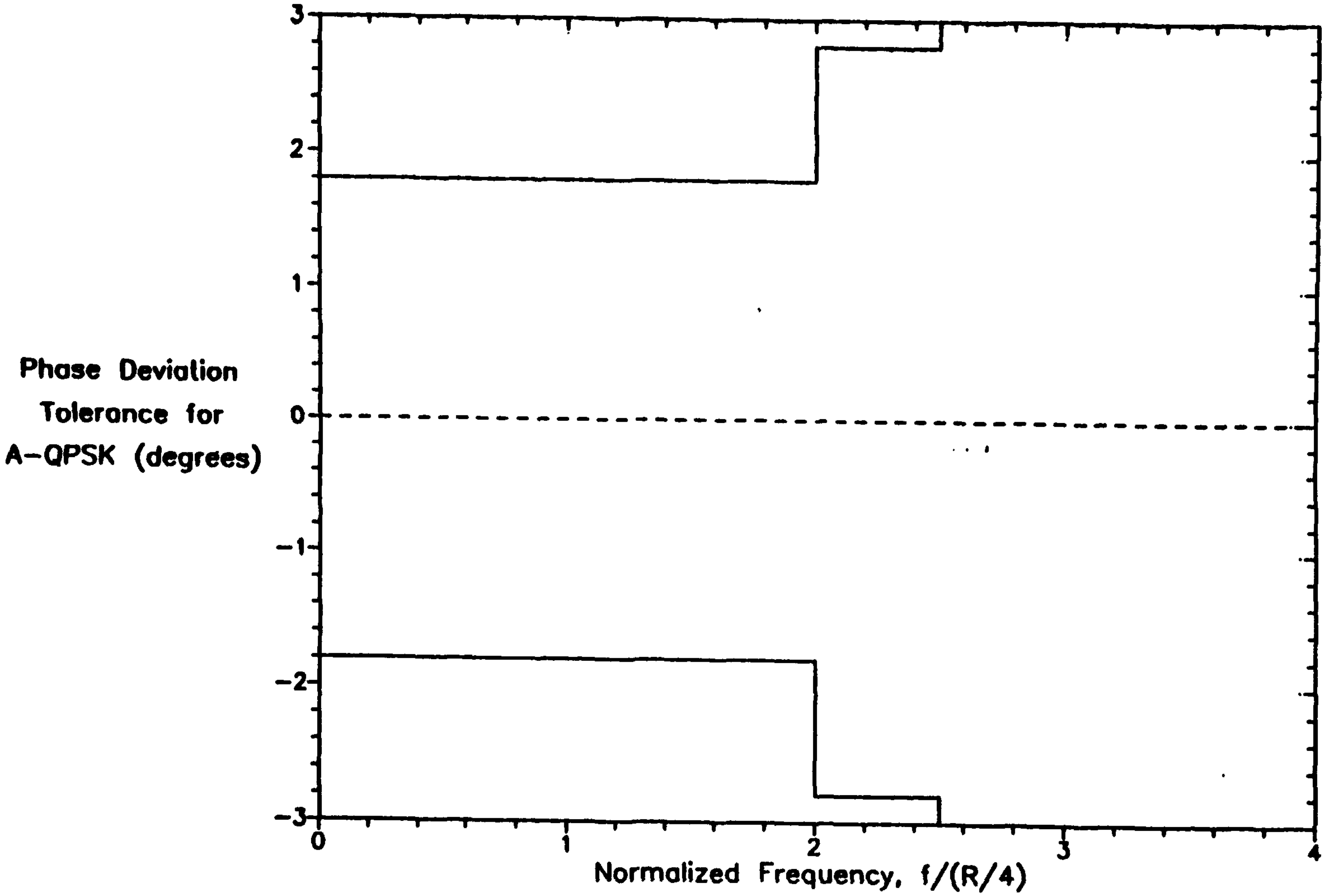


Figure 8-3. Phase Mask Requirement for A-QPSK Signals

MSK Signaling

MSK [5] can be viewed as a special case of OQPSK with half sinusoidal pulses which ensures the constant envelope property but completely avoids the possibility of phase step changes every T seconds, i.e., the signal phase is continuous and the spectral sidelobes drop at a faster rate than with OQPSK. Thus, except for the main lobe, MSK is spectrally more compact than QPSK or OQPSK. The continuous phase nature of MSK signals makes it highly desirable for use with high-power nonlinear transmitters. However, the wider mainlobe suggests that MSK may not be the preferred modulation technique for narrowband links.

The MSK waveform can also be regarded as a continuous phase FSK waveform with signaling frequencies $f_0 + 1/4T$ and $f_0 - 1/4T$, i.e., BFSK with a modulation index of 0.5. MSK can be detected coherently or, since MSK is a type of FSK, it can also be noncoherently detected using a limiter/discriminator. This permits inexpensive demodulation of MSK when the value of received E_b/N_0 permits.

GMSK Signaling

GMSK [6] is obtained by manipulating the output power spectrum of MSK with a premodulation pulse shaping lowpass filter. The lowpass filter is a Gaussian shaped lowpass filter and it has the following properties: 1) narrow bandwidth and sharp cutoff which suppress the high frequency components for narrowband radio applications; 2) low overshoot impulse response which limits excessive instantaneous frequency deviation; and 3) preservation of the filter output pulse area which corresponds to the MSK phase shift of $\pm\pi/2$ with each data bit. The last feature permits coherent detection as in MSK. The spectral compactness of the GMSK depends on the bandwidth of the premodulation Gaussian filter. Values of 0.25 and 0.5 for the normalized 3 dB bandwidth are commonly used for the majority of GMSK implementation. A more efficient spectrum can be obtained by decreasing the filter bandwidth but this results in a longer pulse duration and causes additional ISI that degrades BER performance.

GMSK modulation is used in the pan-European digital cellular telephone, called GSM [7] (Special Mobile Group).

$\pi/4$-QPSK Signaling

In $\pi/4$-QPSK [8], there are two signal sets each consisting of 4 signals spaced by $\pi/2$. The two sets of signals differ by $\pi/4$ with respect to each other as shown in figure 8-1. Differential encoding is achieved by selecting one phase from one set in a given signaling interval, and selecting a phase from the other set in the next interval. Therefore, a relative phase shift of $\pm\pi/4$ or $\pm3\pi/4$ occurs between two successive symbol times.

$\pi/4$-QPSK signals can be detected by a phase comparison detector, or by a limiter/frequency discriminator and an integrate and dump filter. This simplifies the receiver implementation. The envelope of $\pi/4$-QPSK signal has a maximum of 135° phase shift (no zero crossing), and therefore has an advantage, as does OQPSK, of less spectrum spreading due to hardware nonlinearities.

Differential $\pi/4$-QPSK is used in the U. S. digital cellular radio telephone system [9] to provide a 48.6 kb/s channel data rate in a 30 kHz channel time division multiple access (TDMA) system.

M-QAM and M-OQAM Signaling

The principle of quadrature modulation used with QPSK, as well as its constant envelope derivatives discussed briefly above, can be generalized to include amplitude as well as phase modulation to obtain the M-ary QAM signals. M-QAM and M-OQAM (with $M>4$) are of interest because of their high bandwidth efficiency, but the constant envelope property is lost. M-QAM/M-OQAM consists of two independently amplitude-modulated carriers in quadrature. Each block of $k=\log_2 M$ bits (k is assumed to be even so M is an even power of two) can be split into two k/2-bit blocks which use k/2-bit digital-to-analog converters to provide the required amplitudes for the quadrature carriers. At the receiver, each of the two quadrature signals is independently detected using $\sqrt{M}$ matched filters. QAM can be viewed as amplitude shift keying in two dimensions or a hybrid of amplitude shift keying and phase shift keying. For the special case of two antipodal levels on each quadrature channel (a shaping filter can also be applied as with A-QPSK), 4-QAM is identical to QPSK and 4-OQAM is identical to OQPSK. Higher level QAM signal constellations can be created by

choosing an even number of symmetrical levels, e.g., $\pm a$ and $\pm 3a$, on each of the two quadrature channels (again, shaping filters can be used); this yields 16-QAM, or 16-OQAM if the I-and Q-channel bit streams are offset.

4-OQAM is adapted by ARINC for a VHF digital radio [10] and 16-QAM has applications in the high data rate satellite broadcast and communication links [11] and microwave data radios [12].

16-QAM and 16-OQAM are distinctly different from 16-PSK (cf., figure 8-1). Due to their nonconstant envelope nature, 16-QAM and 16-OQAM suffer greater BER performance degradations in a multipath fading environment than constant envelope modulation schemes. QAM and OQAM signals require coherent detection.

8.4 TRADEOFFS

System tradeoffs are fundamental to all digital communications designs. The issues that should be addressed when evaluating a digital modulation technique for VHF A/G communications were delineated in the last section. Tradeoffs among the digital modulations for each issue are discussed here.

Bit-error-rate performance

The BER performance of a digital modulation technique depends on the detailed channel characteristics and the detection method (coherent and noncoherent detection) employed. Coherent detection exploits knowledge of the carrier's phase, thereby providing the optimal error performance attainable with a digital modulation. When, however, it is impractical or too costly to have knowledge of the carrier phase at the receiver, noncoherent or suboptimal detection can be used.

The BER for all digital modulation techniques decrease monotonically with increasing E_b/N_0 for both coherent and noncoherent detections. Selecting a digital modulation technique based on BER performance is to select the modulation/demodulation scheme which has the lowest E_b/N_0 while still meeting the required BER to achieve the given level of message reliability or speech intelligibility. In the ATS or AOC situation, for example, a packet of 1024 bits may be required to have an undetected packet error rate equal to 10^{-6}. This might translate into a 10^{-5} information BER when a cyclic redundancy check (CRC) code and an automatic repeat request (ARQ) retransmission protocol are used in the data transmissions protocol. For digital speech, a BER=10^{-2} is sufficient for many state-of-art vocoders to produce good quality voice.

Table 8-1 shows the E_b/N_0 values needed to achieve certain representative values of BER for various digital modulation schemes and detection methods in an additive white Gaussian noise (AWGN) channel. Although multipath fading can exist in the ATC A/G communication channel, Table 8-1 still provides an indication of relative power efficiency of the various modulation techniques considered in the benign, and perhaps more common, environment.

For any value of E_b/N_0, coherent BPSK requires the smallest E_b/N_0 (9.6 dB for BER= 10^{-5} and 10.5 dB for BER=10^{-6}) of any binary or 4-ary scheme. Although not shown in this table, BPSK and DPSK require an E_b/N_0 that is 3 dB less than the corresponding values for coherent BFSK and noncoherent BFSK, respectively, to realize the same BER. When detected in phase quadrature, MSK, π/4-QPSK, QPSK and 4-QAM have the same BER performance as BPSK. Offset keyed modulations have the same BER performance as their synchronized/aligned counterparts. In the M-ary PSK case, E_b/N_0 increases as M increases for a fixed BER. On the other hand, E_b/N_0 decreases as M increases for a fixed BER in orthogonal M-ary FSK. In other words, M-ary signaling produces improved BER performance with orthogonal FSK signaling and degraded BER performance with M-ary PSK signaling. However, in the case of orthogonal M-ary FSK, BER performance

Table 8-1. Quantitative Comparison of Several Advanced Digital Modulations (Additive White Gaussian Noise)

Digital Modulation Technique	Spectral Compactness* Normalized to Rb**	Bandwidth Efficiency (b/s/Hz)	Power Efficiency Eb/No(dB) for BER			Detection Scheme
			10^{-4}	10^{-5}	10^{-6}	
MSK	0.62	1.6	8.4	9.6	10.5	coherent
GMSK (BT=0.25)	0.5 - 0.62	2.0 - 1.6	10.0	11.2	12.0	coherent
A-QPSK	≤0.5	≥2.0	9.2	10.4	11.3	coherent
π/4-QPSK	0.5	2.0	8.4	9.6	10.5	coherent
8-FSK	1.33	0.75	7.3	8.3	9.2	coherent
16-PSK	0.25	4.0	16.0	17.5	18.8	coherent
8-PSK	0.33	3.0	11.7	13.0	13.9	coherent
BPSK	1.0	1.0	8.4	9.6	10.5	coherent
4-OQAM	0.5	2.0	8.4	9.6	10.5	coherent
16-OQAM	0.25	4.0	12.2	13.4	14.4	coherent
16-QAM	0.25	4.0	12.2	13.4	14.4	coherent

Table 8-1. Quantitative Comparison of Several Advanced Digital Modulations (Concluded) (Additive White Gaussian Noise)

Digital Modulation Technique	Spectral Compactness* Normalized to Rb**	Bandwidth Efficiency (b/s/Hz)	Power Efficiency Eb/No(dB) for BER 10^{-4}	10^{-5}	10^{-6}	Detection Scheme
MSK	0.62	1.6	11.5	12.7	13.6	Limiter/ Descriminator
GMSK (BT=0.25)***	0.5 - 0.62	2.0 - 1.6	13.6	15.0	15.9	Limiter/ Descriminator
A-QPSK	≤0.5	≥2.0	11.6	12.8	13.7	Phase comparison
π/4-QPSK	0.5	2.0	10.8	12.0	12.9	Phase comparison
8-FSK	2.67	0.38	8.2	9.1	9.9	Non-coherent
16-PSK	0.25	4.0	19.2	20.3	21.1	Phase comparison
8-PSK	0.33	3.0	14.6	15.8	17.6	Phase comparison
BPSK	1.0	1.0	9.3	10.2	11.2	Phase comparison

* All based on the noise bandwidth defined by bo = $[1/G(0)] \int_{-\infty}^{\infty} G(f)\, df$, where G(f)is the power spectral density of the modulation waveform, except for 8-FSK which is based on the frequency separation for orthogonality.

** Rb = bit rate.

*** B is the 3 dB bandwidth of the Gaussian filter and T is the bit duration.

improvement is achieved at the expense of bandwidth expansion; while in the case of M-PSK enhanced bandwidth performance is achieved at the expense of BER performance (see bandwidth efficiency section).

The BER performance of GMSK modulation with coherent detection degrades from the antipodal transmission due to the ISI effect of the Gaussian premodulation lowpass filter. The degradation depends on the 3 dB bandwidth of the Gaussian premodulation filter. As shown in the Table 8-1, the E_b/N_0 for GMSK with a normalized (with respect to the data rate) 3 dB filter bandwidth of 0.25 degrades approximately 1.5 dB from that of MSK.

A-QPSK is a filtered OQPSK signal. As a result of filtering, an E_b/N_0 power loss degradation is approximately 0.8 dB. $\pi/4$-QPSK's BER performance is the same as that of QPSK if detected coherently or the same as that of DQPSK if detected differentially with the phase comparison method.

16-QAM outperforms 16-PSK (by approximately 4.1 dB at BER=10^{-5}) due to the fact that the distance between the message points of 16-PSK is smaller than the distance between the message points of 16-QAM as shown in figure 8-1. Again, 16-QAM and 16-OQAM have the same BER performance.

Spectral Compactness

The spectrum compactness of a digital modulation technique is determined by its power spectral density function which in turn depends on the pulse shape transmitted and the pulse duration.

For modulation techniques such BPSK, QPSK and MSK, the spectrum of the (random) modulated data stream is just the spectrum of an individual data pulse, and that expression is mathematically simple. The BPSK data pulse has a rectangular shape resulting in a power spectral density that has the following form [5]:

$$S(f) = P \cdot T \left[\frac{\sin \pi(f-f_o)T}{\pi(f-f_o)T} \right]^2 \qquad (2)$$

where P is the average power in the modulated waveform, T is the pulse duration and f_o is the carrier frequency. Note that the power spectral density function falls off as the square of the frequency.

QPSK and OQPSK pulses also have rectangular shape except that the pulse duration is twice as long. The power spectral density of QPSK and OQPSK is given by [5]:

$$S(f) = 2\,P\cdot T \left[\frac{\sin 2\pi(f-f_o)T}{2\pi(f-f_o)T} \right]^2 \tag{3}$$

As shown in figure 8-4 the power spectral density function falls off slightly faster than that of BPSK since the pulse duration is doubled.

MSK pulses are half cosine-shaped. As a result, the waveform is continuous at bit transition times and the power spectral density falls off as the fourth power of frequency. The typical power spectral density of MSK is given by [5]:

$$S(f) = \frac{16\,P\cdot T}{\pi^2} \left[\frac{\cos 2\pi(f-f_o)T}{1 - 16\,(f-f_o)^2\,T^2} \right]^2 \tag{4}$$

In the usual BFSK waveform, two transmitted frequencies differ by an amount equal to the bit rate, 1/T, and their arithmetic mean equals the carrier frequency f_o. The power spectrum of the BFSK signal contains two discrete frequency components and one continuous component [13]. The two discrete components are located at the two transmitted frequencies with their average powers adding up to one-half the total power of the BFSK signal. The continuous component ultimately falls off as the fourth power of the frequency provided the BFSK signal is phase continuous. For a BFSK signal that is not phase continuous, the power spectral density ultimately falls off as the square of the frequency. It then follows that an FSK signal with continuous phase does not produce as much interference outside the signal band of interest as an FSK signal with discontinuous phase.

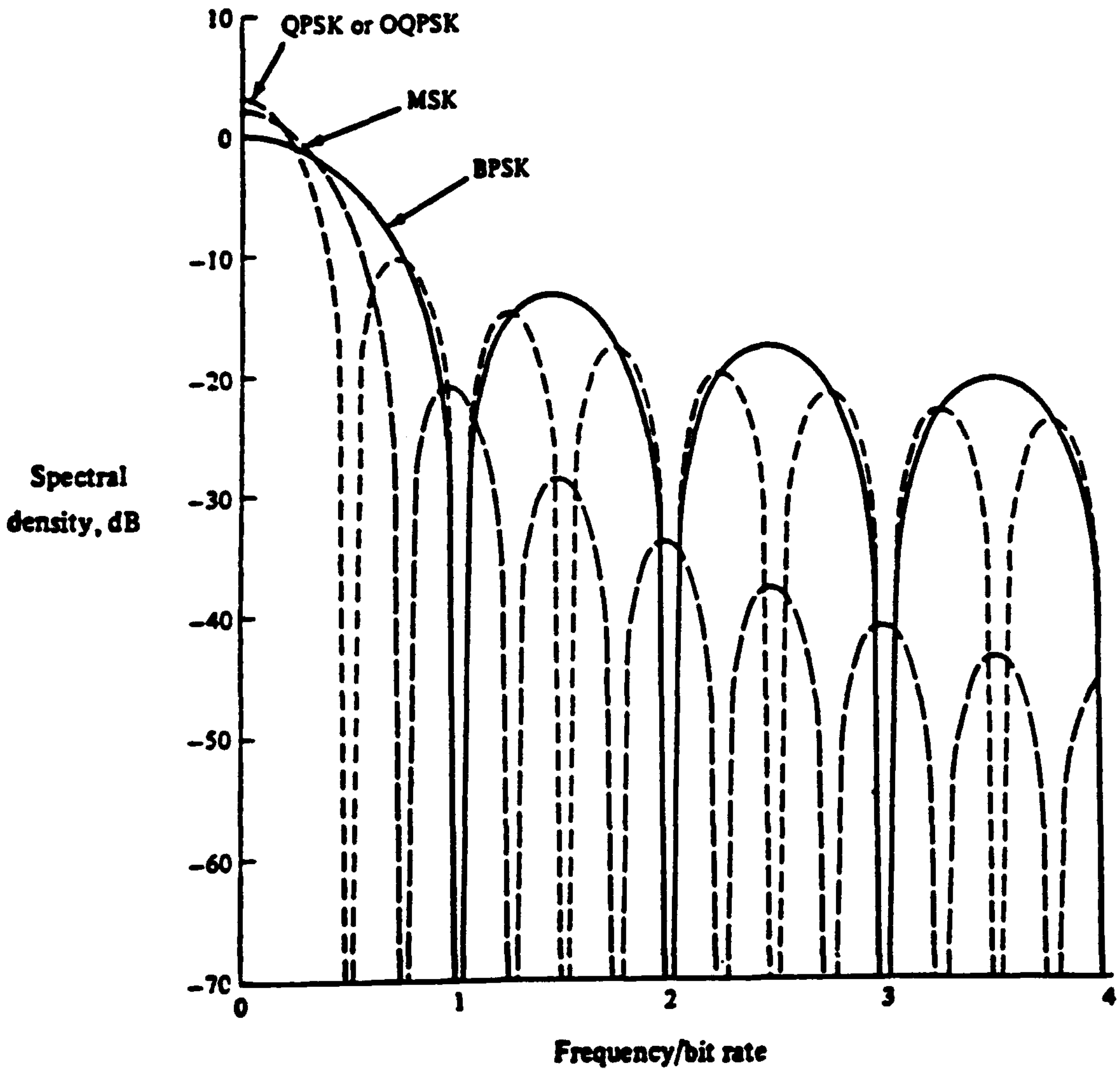


Figure 8-4. Baseband Equivalent Power Spectra
(Baseband bandwidth used)

Figures 8-4 and 8-5 respectively show the power spectral density and its falloff rate for unfiltered BPSK, QPSK, OQPSK and MSK signals. These represent the basic spectra of the digital modulations considered for VHF A/G communications. The difference in the falloff rates of these spectra can be attributed to the smoothness of the phase transition at the bit transition point. The smoother the phase change of the transmitted pulses during the bit transition, the faster is the drop of the spectral tails to zero. Thus, BFSK with continuous phase has lower sidelobes than BPSK. The BPSK waveform has a constant envelope but has an abrupt phase change at the bit transitions; if the modulating data stream were to consist of alternating zeros and ones, there would be such an abrupt change at each transition. Due to the abrupt change at bit transitions, the power spectral density has much higher sidelobes and the falloff is much slower than MSK. QPSK and OQPSK also have abrupt phase changes at the bit transitions. However, because the pulse duration is doubled that of BPSK, the power spectral density of QPSK and OQPSK has lower sidelobes and falls off faster than BPSK. MSK has a smoother transition and thus has lower sidelobes than QPSK and OQPSK. As we see from figure 4, MSK is spectrally more efficient than QPSK and OQPSK overall. However, the MSK spectrum has a wider main lobe than QPSK and OQPSK. Table 8-2 summarizes the noise, the null-to-null, half-power and 99% power containment bandwidths required for these modulation techniques [14]. The values are normalized to data rate, 1/T.

For basic modulation techniques such as BPSK, QPSK and MSK, the spectrum of the (random) modulated data stream is just the spectrum of the individual data pulse, and that expression is mathematically simple as shown earlier. The spectra of higher order phase and amplitude shift keying signals, such as M-PSK (8-PSK and 16-PSK), 8-FSK and 16-QAM (or 16-OQAM), take the same form as BPSK, BFSK and QPSK respectively, but with the frequency variable rescaled to reflect the bandwidth efficiency.

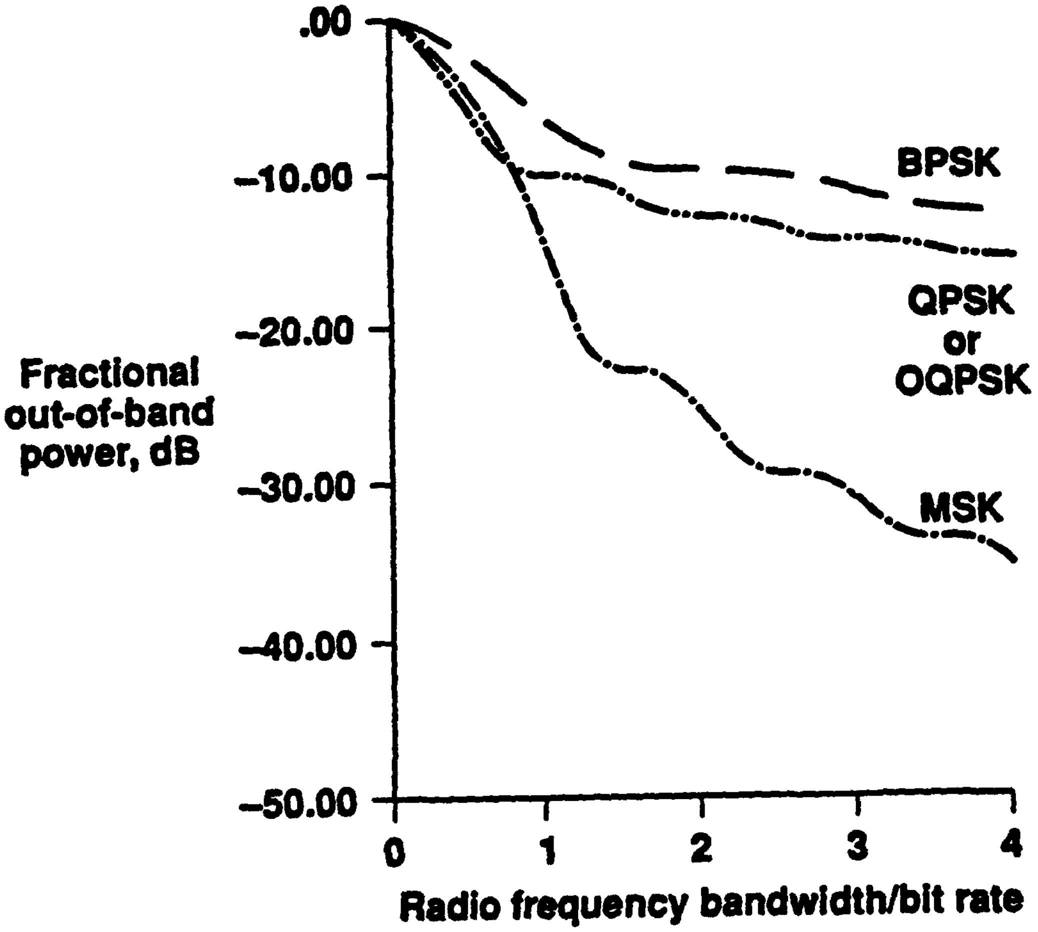


Figure 8-5. Factional Out-of-Band Power for Various Modulation Schemes

A-QPSK is a filtered OQPSK signal with the sidelobes reduced to at least 40 dB from the peak. The spectrum falloff is faster than that of OQPSK. $\pi/4$-QPSK is two $\pi/4$ offset QPSK signals and has the same spectral characteristics as that of QPSK. GMSK incorporates a Gaussian-shaped low-pass filter to smooth the phase transitions of the MSK signal, thus, yielding a narrower and faster falloff spectrum than MSK. The roll-off of the filter is determined by its 3 dB bandwidth. The choice of a 3 dB bandwidth value is a trade-off between the spectral efficiency and the receiver complexity. A small value introduces a high degree of ISI, which requires a complex receiver to demodulate. For a typical 3 dB bandwidth equal to 0.25 (normalized to 1/T), GMSK has a 99% power containment bandwidth of 0.86 compared to 1.18 of MSK as shown in Table 8-2.

8-FSK utilizes channel bandwidth inefficiently. This is because only a fraction of the bandwidth is occupied by the transmission of one of the 8-orthogonal carriers (tones) during any symbol period. Consequently, the use of 8-FSK is limited to situations where bandwidth conservation is not of primary concern. 8-PSK and 16-PSK provide a more efficient utilization of channel bandwidth than their FSK counterparts. For PSK signals, the bandwidth efficiency increases with increasing M. However, the BER performance decreases with increasing M. Channel utilization may be improved further by the use of hybrid amplitude and phase modulation such as 16-QAM or 16-OQAM. This is due to the ability to independently modulate the I- and Q- components of the signal. However, 16-QAM and 16-OQAM require the channel to be fairly linear since they have a nonconstant envelope. By contrast, 16-PSK is more resistant to the amplitude nonlinearities in the channel due to its constant signal envelope, but not to the phase distortions resulting from nonlinearities.

The relative spectral compactness of various modulations for VHF A/G systems is also shown in Table 8-1. The values are based on the noise bandwidth of the modulated signal (except for FSK signals) and they are normalized with respect to the information rate. The noise bandwidth, b_o, of a digital modulated signal with a power spectral density S(f) is defined as the bandwidth of an equivalent signal that has a constant power spectral density $S(f_o)$ and the same total power as the actual signal, i.e.,

Table 8-2. Various Bandwidths for Some Digital Modulation Schemes Normalized to Data Rate, 1/T)

Modulation Schemes	Null-toNull Bandwidth	Noise Bandwidth	Half-Power Bandwidth	99% Power Containment Bandwidth
BPSK	2.00	1.00	0.88	20.56
QPSK/OQPSK	1.00	0.50	0.44	10.28
MSK	1.50	0.62	0.59	1.18
GMSK	1.50	-	0.58	0.86

$$b_o = \frac{1}{S(f_o)} \int_{-\infty}^{+\infty} S(f)\, df \qquad (5)$$

where $S(f_0)$ is the power spectral density at the carrier frequency. The noise bandwidth is often used in connection with passing AWGN through filters. The noise bandwidth helps evaluate data link performance in the face of ACI and CCI.

ACI is characterized as unwanted signals from other frequency channels "spilling over" or injecting energy into the channel of interest. The severity of ACI and therefore the choice of channel bandwidth is mainly determined by the spectral roll-off of the modulated signal as described earlier. In general, there will be less ACI for a system with a modulated signal having a fast spectral falloff. ACI can also be reduced by geographically separating the adjacent channels. For a system with a BER=10^{-5}, the E_b/N_0 (from Table 8-1) is 9.6 dB for 4-OQAM and 4-QAM, 13.0 dB for 8-PSK, 13.4 dB for 16-OQAM and 16-QAM and 17.5 dB for 16-PSK when there is no interference. In the presence of ACI, additional E_b/N_0 is required to obtain the same BER and the additional E_b/N_0 depends on the modulation scheme, the channel separation and the geographic frequency assignment. Among the QAM and PSK signals the ranking of systems according to the BER performance is [15]: 4-QAM and 4-OQAM, 16-QAM and 16-OQAM, 8-PSK, 16-PSK, if only AWGN and ACI are taken into account.

Another problem associated with the geographically distributed transmitters and receivers in the ATC environment is the near-far problem. The near-far problem comes about when the desired transmitter is further from the receiver than some of the undesired transmitters. Thus the amplitude of the undesired signal can be much higher than the desired signal. Among the modulation schemes considered, 16-QAM and 16-OQAM probably are the most sensitive ones to the near-far problem. However, if all ATC transmitting channels are assigned with different frequencies, the receiver can reject the undesired adjacent channel signals and thus the near-far problem is less significant.

Bandwidth Efficiency, R/W

The bandwidth, W, required to transmit a digital signal at a rate, R=1/T, is simply the bandwidth of its equivalent low-pass signal pulse which depends on its detailed characteristics. The data rate is well defined. Unfortunately, however, there is no universally satisfying definition for the bandwidth, W. This means that the bandwidth efficiency of a digital modulation scheme depends on the particular definition adopted for the bandwidth of the modulated signal. The bandwidth of a digital signal can be measured in several different ways, such as null-to-null bandwidth, noise bandwidth, half-power bandwidth, 50 dB bandwidth, 99% power containment bandwidth, etc.

The null-to-null bandwidth encompasses the main lobe of the power spectrum of the signal. The noise bandwidth specifies the equivalent bandwidth of a rectangular power spectrum that would have the same total power. The noise bandwidth has been discussed in the last section. Other bandwidth definitions are related to bounding power spectral density by specifying that everywhere outside the bandwidth the power spectral density must have fallen at least to a certain stated level below that found at the band center.

For our purpose, we assume that the digital signal is a lowpass pulse of duration, T, and its bandwidth (except for FSK signals), W, is approximated by its noise bandwidth which is given in Table 8-2. The bandwidths for the FSK signals shown in Table 8-1 are based on the frequency separation to maintain the orthogonality as discussed earlier. Note that bandwidth values in Table 8-1 are normalized to the information bit rate, $R_b = \log_2 M / T$. For a BPSK signal, the required normalized signaling bandwidth, WT, is 1. The bandwidth efficiency, R/W, in b/s/Hz of various modulation techniques is tabulated in Table 8-1.

From Table 8-1, we see that an in-phase and quadrature scheme is preferred, since, to the first order, this doubles the data rate over just using the in-phase channel, such as in BPSK or BFSK. Beyond that M-ary PSK or M-ary QAM signaling provides even higher bandwidth efficiencies. However, a higher bandwidth efficiency is accompanied by a higher E_b/N_0 to achieve a certain BER performance. In other words, in this case, we are trading power efficiency for bandwidth efficiency. On the other hand, M-ary FSK signaling does not provide high bandwidth efficiencies; instead, it trades bandwidth efficiency for power efficiency. An M-ary FSK signal consists of an orthogonal set of M frequency-shifted

signals. The adjacent signals need be separated from each other by a frequency difference 1/2T in order to maintain orthogonality, when the orthogonal signals are detected coherently. The frequency difference is 1/T when the signals are detected noncoherently.

Effect of Nonlinearities

A nonlinear channel will produce extraneous sidebands when passing a signal with amplitude fluctuations. Such sidebands may cause interference, such as CCI and ACI, with its own and other communication systems. This inevitably causes a degradation in BER performance or frequency assignment restrictions that reduce system capacity.

All constant envelope modulation techniques are not equally insensitive to nonlinearities. For example, in QPSK, due to the alignment of the I and Q data streams, the carrier phase can change only once every 2T. The maximum phase change is 180° when both I and Q data stream change signs. The BPSK carrier phase can change at every bit transition time. The maximum phase change is also 180°. When a constant envelope QPSK or BPSK modulated signal undergoes filtering to reduce the spectral side lobes, the resulting waveform will no longer have a constant envelope and in fact, the occasional 180° phase shifts will cause the envelope to go to zero momentarily. When the band-limited QPSK or BPSK signal goes through a nonlinear amplifier, the constant envelope will tend to be restored. However, at the same time, all of the undesirable frequency sidelobes, which can interfere with nearby channels and other communications, are also restored.

In OQPSK, the I and Q data streams are staggered and thus do not change states simultaneously. A change of carrier phase by 180° is eliminated. The maximum change is ±90° every T seconds. When an OQPSK signal undergoes bandlimiting, the resulting intersymbol interference causes the envelop to droop slightly in the regions of ±90° phase transitions. Since the phase transitions of 180° have been avoided in OQPSK, the envelope will not go to zero as it does with QPSK. When bandlimited OQPSK goes through a nonlinear amplifier, the envelope droop is removed and some sidelobes are restored; however, the out-of-band interference is much reduced. This is the main advantage of OQPSK over QPSK. Therefore, OQPSK (or 4-OQAM), and its filtered version, A-QPSK, are favored over QPSK modulation. The $\pi/4$-QPSK would then be a compromise between QPSK and OQPSK. The $\pi/4$-QPSK signal also has a constant envelope. Its maximum phase

change is $3\pi/4$, hence, the spectral spread caused by nonlinearities also would not be as large as in QPSK.

Due to the half cosine weighting on their rectangular pulses, MSK and GMSK signals have continuous phase transitions at the bit transition times ensuring that the amplitude of the modulated carriers are essentially constant even after bandpass filtering. Hence there is little spectral spreading even if nonlinear devices are used. This greatly reduces the ACI problem.

In summary, the spreading of the spectrum of bandlimited BPSK, QPSK and M-PSK signals occurs when they are passed through a nonlinear device. This effect is associated with a sudden and major change in the amplitude of the carrier. OQPSK, 4-OQAM, A-QPSK and $\pi/4$-QPSK lessen these amplitude changes and thus reduce spectrum spreading. The step changes in amplitude are avoided completely in MSK and GMSK signals thus minimizing the nonlinearity effect .

System Robustness

Multipath exists when there is more than one transmission path between transmitter and receiver. In the VHF ATC A/G radio environment, signal transmission between the ground station and the aircraft may occur via random multiple propagation routes in addition to the primary direct path. Consequently, fast and slow multipath fading may appear on the received direct path signals and degrade the signal BER performance.

Slow fading is caused by such physical phenomena as specular earth reflection, foliage attenuation, atmospheric refraction, atmospheric multipath, and airborne antenna shadowing. Slow fading can last for seconds or even minutes. Fast fading is caused by such physical phenomena as diffuse earth reflection, particularly from buildings, airframe reflections and diffraction, and rotor-wing aircraft blade reflections.

Theoretically, multipath fading can be modeled by assuming Rayleigh fading statistics or Rician fading statistics. In Rayleigh fading, there are a large number of signals which vary independently, with amplitudes roughly equal and phases approximately random; while in Rician fading statistics, in addition to the desired signal from the direct path, there is a

strong interference arising from another relatively stable path. In ATC A/G communication channels, the secondary transmission paths from buildings, earth reflection, foliage, etc., can be characterized by the summation of several delayed and attenuated replicas of the desired signal. Thus, the channel fading statistics may lean toward the Rician fading channel instead of the Rayleigh fading channel. Unfortunately the quantitative results on the BER performance of various modulations under Rician fading channels are not available. Nevertheless, to indicate that the BER performance degradation due to multipath fading could be severe, we summarize the BER performance under Rayleigh fading in Table 8-3.

Table 8-3 demonstrates that selecting a digital modulation for the ATC channel based on its BER performance in an AWGN channel and assuming a constant fading margin, (e.g., 10 dB is commonly used) is not a good approach. From Table 8-3, we see that Rayleigh fading could cost a few tens dB of E_b/N_0 degradation. In the ATC channel, the fading degradation is not as severe as that in the Rayleigh fading channels. The required fading margin depends on the type of modulation, the required information BER, the detection method and the channel characteristics. BER improvement techniques such as error correcting codes and/or time, frequency, or space diversity may be required for effective data transmission in the ATC VHF channels to reduce the degradation. These will be addressed in a separate decision tree memorandum under branchpoint 4.1.2.2, Coding Technique.

Implementation Complexity

Many of the modulation schemes have a number of alternative implementation options. In general, for a given system bandwidth and BER requirement, the complexity and cost of the modulation scheme increases as the bandwidth efficiency of the modulation scheme increases. The relationship between complexity and bandwidth efficiency does not appear to be linear.

Table 8-3. Quantitative Comparison of Several Advanced Digital Modulations (Rayleigh Fading Channel)

Digital Modulation Technique	Spectral Compactness* Normalized to Rb**	Bandwidth Efficiency (b/s/Hz)	Power Efficiency Eb/No(dB) for BER			Detection Scheme
			10^{-2}	10^{-4}	10^{-6}	
MSK	0.62	1.6	14.0	34.0	54.4	coherent
GMSK (BT=0.25)***	0.5 - 0.62	2.0 - 1.6	17.5	36.0	55.0	coherent
A-QPSK	≤ 0.5	≥ 2.0	14.3	33.8	53.8	coherent
p/4-QPSK	0.5	2.0	13.5	33.0	53.0	coherent
8-FSK	1.33	0.75	15.0	36.0	60.5	coherent
16-PSK	0.25	4.0	21.0	-	-	coherent
8-PSK	0.33	3.0	16.5	35.0	54.0	coherent
BPSK	1.0	1.0	14.0	34.0	54.4	coherent
4-OQAM	0.5	2.0	13.5	33.0	53.0	coherent
16-OQAM	0.25	4.0	-	-	-	coherent
16-QAM	0.25	4.0	-	-	-	coherent

Table 8-3. Quantitative Comparison of Several Advanced Digital Modulations (Concluded) (Rayleigh Fading Channel)

Digital Modulation Technique	Spectral Compactness* Normalized to Rb**	Bandwidth Efficiency (b/s/Hz)	Power Efficiency Eb/No(dB) for BER 10^{-2}	10^{-4}	10^{-6}	Detection Scheme
MSK	0.62	1.6	17.0	37.0	57.0	Limiter/ Discriminator
GMSK (BT=0.25)***	0.5 - 0.62	2.0 - 1.6	21.0	41.0	62.0	Limiter/ Discriminator
A-QPSK	≤ 0.5	≥ 2.0	20.8	37.8	57.8	Phase comparison
π/4-QPSK	0.5	2.0	14.5	37.0	57.0	Phase comparison
8-FSK	2.67	0.38	18.0	39.0	63.5	Noncoherent
16-PSK	0.25	4.0	24.0	-	-	Phase comparison
8-PSK	0.33	3.0	19.4	38.5	58.4	Phase comparison
BPSK	1.0	1.0	17.0	37.0	57.0	Phase comparison

* All based on the noise bandwidth defined by $b_o = [1/G(0)] \int_{-\infty}^{\infty} G(f)\, df$, where G(f)is the power spectral density of the modulation waveform, except for 8-FSK which is based on the frequency separation for orthogonality.

** Rb = bit rate.

*** B is the 3 dB bandwidth of the Gaussian filter and T is the bit duration.

Coherent detection requires a replica of carrier frequency and phase. Carrier tracking and phase jitter compensation, especially in a multipath fading environment, could be expensive, requiring a complex system design. Noncoherent or phase comparison detection techniques are often preferable since they are generally simpler and less costly than their coherent counterparts. The price paid for noncoherent or phase comparison detection is that a higher E_b/N_0 is required to achieve the same BER performance. When received E_b/N_0 permits, FSK, MSK and GMSK modulated signals can be detected by a FM-discriminator which is relatively easy to implement. On the other end, the highly bandwidth efficient 16-QAM and 16-OQAM techniques require a linear channel. The design of an efficient linear amplifier is much more complex than the design of an efficient nonlinear amplifier. Thus modulations like 16-QAM and 16-OQAM are more difficult to implement.

8.5 IMPACT/IMPORTANCE

Choosing a digital modulation would require replacing the current analog AM A/G radio in the VHF 118-137 MHz band by a brand-new digital radio. Technology trends clearly indicate that the future is with digital modulation.

In particular, digital modulation has many important features which current analog modulation cannot provide. First, digital modulations provide a much greater information transmission rate than that of current AM systems. Digital techniques also can provide more reliable data transmissions and alternate routing/networking capabilities. The current ACARS data capabilities, implemented by 2400 b/s or 4800 b/s MSK signalling in the current VHF analog radio, can be improved upon with a higher data rate digital VHF radio. The digital modulation with higher bandwidth efficiency, such as 4-OQAM, 16-OQAM, etc., can also reduce the spectrum congestion problem in the ATC A/G environment.

Resistance to fading is another important reason to use digital modulation. Data transmission in the ATC channel can suffer system performance degradation from slow and fast multipath fading. This problem can be controlled with digital modulations by using error correcting codes combined with interleaving and ARQ techniques. However, the information rate (and bandwidth efficiency) is reduced when a strong error correcting code is used. These techniques will be discussed in a separate decision tree memorandum under branchpoint 4.1.2.2, Coding Technique.

Another important consideration in choosing between digital and analog modulation is whether or not voice and data communications are to be provided in the same radio. If combining data and voice transmission in a single radio is important, then a digital modulation technique is probably the choice for both functions. This is because of the incompatibility of analog radio with narrow channel spacing for voice with the needs of data transmission which is at much higher transmission rate.

8.6 TRANSITION

Voice in the current AM radio uses DSBTC analog modulation. Data information is transmitted using voice band analog MSK modems at a rate of 2.4 kb/s. The future ATC VHF radio is expected to use some kind of digital voice and data modulation.

For the conversion from analog to digital modulation, it has been pointed out by Wilson [16], that some modifications will require abrupt changes, e.g., analog voice to digital voice, while other modifications will be reached more gradually, e.g., data services.

Modifications for transition will in fact depend on the degree of compatibility of the future improved radio with the current AM radio. Assuming the new radio is required to be backward compatible, a possible approach for transition is as follows. First, while keeping the analog DSBTC voice modulation and MSK data modulation for ACARS functioning, a digital modulation such as 4-OQAM or A-QPSK could be added to provide data service channels for additional functions which the current MSK data modem does not provide. Gradually the digital modulation would take over the data services provided by the MSK modem. Secondly, the modification could include both analog voice and digital voice. Again, gradually the digital voice would take over the voice service.

8.7 CRITERIA FOR DECISION

There are a number of digital modulation/detection techniques that may be suitable for a VHF digital communication system over the ATC channel. Each modulation technique offers tradeoffs of its own. The final choice will depend on the priority given to each of the following goals which may impose conflicting requirements: 1) to maximize the BER performance, i.e., to minimize the E_b/N_0 for a specific BER or to minimize BER for a fixed E_b/N_0; in other words, to minimize the transmitter power for a given link; 2) to minimize the adjacent and co-channel interferences with a signal having good spectral properties so that the interference resistance is maximized; 3) to maximize the information transmission rate; 4) to minimize the effect of nonlinearities from system components; 5) to minimize the BER performance degradation in slow and fast multipath fading channels; and 6) to minimize the design complexity.

The relationship between bandwidth efficiency and power efficiency for a number of digital modulation techniques is shown in figure 8-6. The figure assumes an AWGN channel and a typical BER=10^{-6}. The figure shows that among the PSK modulations, such as BPSK, QPSK, 8-PSK and 16-PSK, etc., QPSK and its derivatives appear to offer the best tradeoff between the power efficiency and the bandwidth efficiency. The power requirements for 8-PSK and 16-PSK are probably excessive for the ATC application. Furthermore, due to possibility of multipath fading, Doppler spread, ACI and CCI, the E_b/N_0 required in practical system applications is typically larger than the theoretical values shown in Tables 8-1 and 8-3. Hence, a power efficient modulation scheme such as QPSK, OQPSK or its derivatives is desired.

Digital modulations with constant, or nearly constant, signal envelopes are desirable in order to avoid the distortion resulting from nonlinearities in hardware implementations.

16-PSK and 16-QAM have been used in point-to-point line of sight (LOS) microwave links due to their high spectral efficiency and relatively good BER performance (when coherent detection is employed). In LOS microwave channels multipath fading is slow and infrequent, but the fading is deep when it occurs. However, Doppler compensation and carrier phase tracking are not much a problem since the transmitter and receiver are stationary. It is easier to implement the coherent detection in LOS microwave links than in

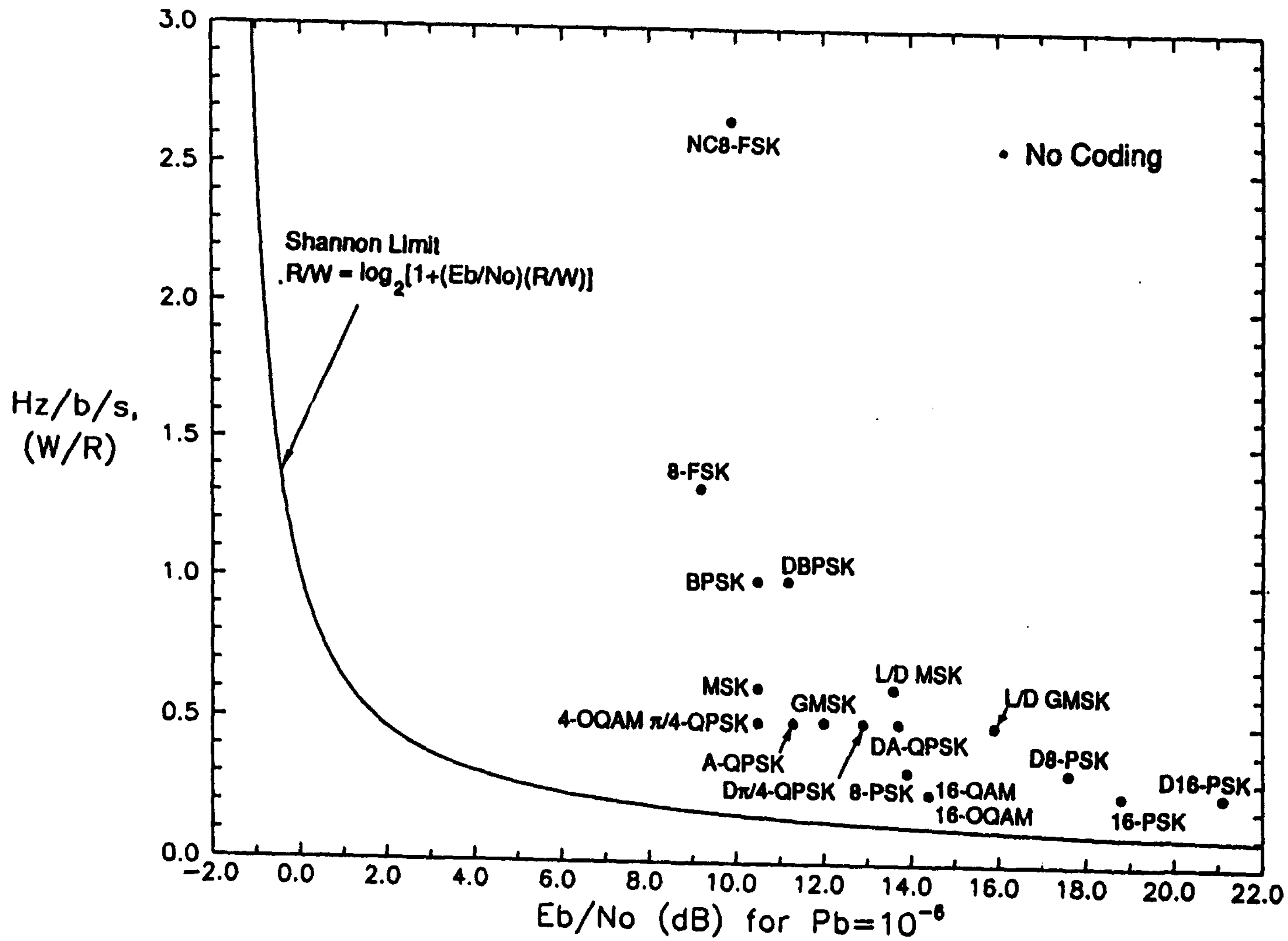


Figure 8-6. Channel Bandwidth-Power Tradeoff for Various Modulation Schemes without Coding

the ATC A/G links. 16-QAM and 16-PSK have similar spectral and bandwidth characteristics. The distance between the signals in 16-PSK is smaller than the distance between the signals in 16-QAM, so 16-QAM outperforms 16-PSK in AWGN. However, the superior performance of 16-QAM is generally realized only if the channel is free from nonlinearities and fading.

As for M-FSK, increasing M results in an increased power efficiency. However, the reduction in transmitted power is achieved at the cost of increasing channel bandwidth which is a precious resource in the ATC systems.

MSK and GMSK have good spectral efficiency and resistance to system nonlinearities. However, their bandwidth efficiency in b/s/Hz for a narrowband system is inferior to QPSK (cf., figure 8-6). Due to the compelling advantage of offset signaling over regular signaling in a nonlinear channel, filtered offset quadrature signaling such as 4-OQAM, A-QPSK and $\pi/4$-QPSK appears to be the digital modulation choice for the near term. Although 16-OQAM requires a linear channel to realize its superior BER performance, the cost to provide phase jitter compensation and tracking may be compensated by its increased bandwidth efficiency compared to the group of OQPSK modulations. It is thus a good candidate for future applications if it can be shown to give satisfactory performance in the presence of multipath fading on the VHF A/G channel..

Use of either coherent or noncoherent detection depends in part on the occurrence of multipath fading in the ATC A/G channel. If coherent detection is to be used, 4-OQAM and A-QPSK have compelling advantages over other modulation techniques due to their considerably better BER performance, high bandwidth efficiency and good resistance to nonlinearity effects and adjacent channel interference. However, severe multipath fading causes irreducible BER performance degradation and carrier recovery would be difficult to obtain. In that case, noncoherent detection would be preferable, and phase comparison $\pi/4$-QPSK and GMSK detected non-coherently with a limiter and discriminator appear to be good candidates.

8.8 CONNECTIVITY/RELATIONSHIP WITH OTHER DECISIONS

This paper has discussed the choices among digital modulations for the VHF A/G system, branchpoint 4.1.2.1 on the decision tree. Section 6 has discussed the choices among the analog modulations, branchpoint 4.1.1 on the decision tree. A level up decision at branchpoint 4.1 of the decision tree, Advanced Modulation/Coding addresses the choice between analog and digital modulation. This has been discussed in section 5. Essentially, analog modulation is a simpler and a nearer-term improvement which addresses primarily the need for more channels, while digital modulation is a more complex longer-term improvement which provides a number of other improvements besides a greater number of channels.

A decision at branchpoint 4.1.2.2 of the decision tree, Coding Technique, will be examined in section 9 in which techniques such as error correcting coding, ARQ and data interleaving will be discussed. These techniques are applied before the modulation and after the demodulation to enhance performance. This section and section 9 on coding techniques form the branchpoint 4.1.2, Digital Modulation /Coding techniques.

SECTION 9

CODING TECHNIQUES

9.1 CONTEXT

A systematic process for examining technical alternatives for improved air/ground communications in air traffic management has been established. A decision tree structure is shown in Appendix A that attempts to organize various alternatives in a top-down hierarchy. This provides a framework for evaluating potential solutions that can be represented by paths through the decision tree.

This paper addresses a particular alternative in the decision tree, namely, branchpoint 4.1.2.2, Coding Technique. The organization of this paper is as follows: Background, Issues, Tradeoffs, Impact/Importance, Transition, Criteria for Decision, and Connectivity/Relationship with Other Decisions. These alternatives are not developed in a vacuum but with sensitivity to the realities of air traffic management.

Cost is treated in this paper only in terms of relative implementation complexity due to limited resources. However, any new system should satisfy Air Traffic Control (ATC) needs in a cost effective way. This means that the one-time cost and life-cycle cost for air/ground (A/G) communications should be minimized. Also, the airborne radio should be affordable by the general aviation community.

9.2 BACKGROUND

Current VHF A/G communications in the 118 - 137 MHz frequency band use the Double Side Band Transmitted Carrier (DSBTC) form of analog modulation, as discussed in section 6. The capabilities of the current system for digital data transmission are limited. The far-term improvement of VHF A/G communications lies with digital modulation, as discussed in section 8. The decision tree outlines a wide range of far-term improvements to ATC VHF A/G communications. Many or most of the improvements require the use of digital modulation. Error control coding techniques combined with digital modulation can provide more reliable data communications, alternate routing/networking and other features that are difficult if not impossible to implement with analog modulation systems. In the

context of error performance, digital modulations deal with transforming waveforms into "better waveforms", to make the detection process less subject to error. Error control techniques deal with transforming data sequences into "better sequences", having structured redundancy. The redundant bits can then be used for the detection and correction of errors.

The main motivation for considering different error control coding techniques for VHF A/G communication is to provide improved BER performance of digital modulations, particularly in the anticipated slow and fast multipath fading environments. Slow fading is caused by such physical phenomena as specular earth reflection, foliage attenuation, atmospheric refraction, atmospheric multipath, and airborne antenna shadowing. Slow fading can last for seconds or even minutes. Fast fading is caused by such physical phenomena as diffuse earth reflection, particularly from buildings, airframe reflections and diffraction, and rotor-wing aircraft blade reflections.

Specifically, the motivations are:

1. To improve the BER performance of digital modulations and provide a more reliable communications link with the available signal power, i.e., to reduce the required E_b/N_0 for a given BER. In the Air Traffic Services (ATS) or Aeronautical Operational Control (AOC) packet data mode, a packet may contain up to 8192 bits and may be required to have an undetected packet error rate equal to 10^{-6}. For digital voice (circuit) mode, a BER $=10^{-2}$ probably is sufficient for a state-of-art vocoder to produce good quality voice.

   Error control techniques are needed when the available E_b/N_0 can not ensure the required BER under the adopted digital modulation scheme, e.g., when it is impractical to boost the transmitted power or to upgrade the hardware to reduce the component losses or to improve the performance. For example, in the packet data mode, if error control techniques are not used, an undetected packet error rate of 10^{-6} approximately corresponds to a BER of 10^{-9} (assuming random errors and 1024 bit packet size) which is too stringent for a practical digital modulation scheme to achieve over the ATC VHF A/G channel.

2. To increase the digital modulated signals' resistance to interferences, such as adjacent channel interference (ACI), co-channel interference (CCI).

3. To increase the robustness of digital modulated signals in the multipath fading communications channel.

An error control coding technique is an essential element for the digital modulation alternatives being considered for the improvement of the ATC communications system. For instance, RTCA Special Committee 172 is considering improvements in VHF A/G communications in the 118 - 137 MHz band to meet future ATS and AOC requirements. Working Group 1 is considering new digital system architectures including various bandwidth efficient digital modulation techniques. Working Group 2 is preparing a Minimum Aviation System Performance Standards (MASPS) to define the signal-in-space characteristics for advanced VHF digital data communications including compatibility with digital voice techniques.

This decision tree memorandum considers various candidate error control coding techniques which could be employed with several of the digital modulations which are candidates to supplant the current analog modulation in the A/G system of the future. First, each of the important issues used in comparing different error control coding techniques is discussed. Then the pros and cons of various error control coding techniques are presented.

9.3 ISSUES

The major issues to be addressed when comparing different error control coding techniques are all related to improved A/G communications reliability. The issues involved are:

1. Performance gain achieved with an error control coding technique.

   Performance gain (or coding gain) is defined as the decrease in signal-to-noise ratio, E_b/N_0, to achieve a given information BER, P_{be}, under the same digital modulation with an error control scheme. It is a measure of the transmission power efficiency improvement.

2. Bandwidth efficiency and data throughput.

The bandwidth efficiency is defined as the number of information bits transmitted per second, R, (in b/s) per unit bandwidth (in Hz) with a specific digital modulation scheme. The bandwidth, W, required to transmit the modulated digital signals at a channel bit rate, R_c, is defined as the bandwidth of the signal's equivalent low-pass signal pulse which depends on its detailed characteristics. For a given digital modulation scheme, R_c/W is fixed, and R_c equals the information data rate, R, when there is no error coding applied. The information data rate contained in the transmission decreases and varies with different error coding techniques applied. As a result, error control coding techniques reduce the bandwidth efficiency in b/s/Hz or data throughput efficiency. The reduction is due to the added redundant bits in the data stream for error detection or correction. Bandwidth efficiency is directly proportional to the code rate of the error correction code, defined as the ratio of information bit rate to transmitted channel bit rate. The data throughput is affected by the data retransmission strategies which will be discussed later.

3. Code robustness, i.e., performance improvement of an error control coded digital modulation scheme versus the same uncoded scheme in both slow and fast fading environments of the ATC VHF A/G communication channels.

4. Design complexity required to implement the coding technique.

Both forward error correction (FEC) and automatic repeat request (ARQ) types of error control techniques should be considered for ATC VHF A/G systems. Both techniques add redundancy to data prior to transmission in order to reduce the effect of errors that occur during transmission. However, the fundamental operational philosophy of the two techniques is quite different. FEC utilizes redundancy so that a decoder can correct the error at the receiver. There does not need to be a return path. ARQ utilizes redundancy to detect errors at a receiver and upon detection, to request a repeat transmission. A return path, and generally a more reliable one, is necessary. Individual error control techniques suitable for ATC VHF A/G communication systems

are now reviewed and discussed in the following subsections to make clear the tradeoff issues.

Forward Error Correcting Codes

The concept of FEC in digital communication is to take the information data, add some redundancy according to a certain algorithm, and transmit both the information data and the redundant data. At the receiver, the noise corrupted signal is processed by a decoder which utilizes the redundancy to reconstruct the correct data. There are many different error-correcting codes that we can use for ATC VHF A/G systems. They can be partitioned into two important categories: block codes and convolutional codes. In this section, various FEC codes are reviewed and the effects of these FEC codings on the four issues listed earlier will be discussed in the tradeoff section.

Block Codes

In the case of block codes, the information bits or message bits are segmented into blocks of k data bits; each block can represent any one of 2^k distinct messages. Block codes are a class of parity check codes. The encoder independently transforms each k-bit data block into a larger (codeword) block of n bits, called channel bits, by an encoding algorithm. The (n - k) bits, which the encoder adds to each data block, are called redundant or parity bits. The code is referred to as an (n,k) code. The ratio, k/n, is called the code rate. The decoder works on a similar basis; each block is individually processed. Note that the decoder has to be informed of the block boundaries by means of the bit and codeword synchronization process in the receiver. The encoding and decoding algorithms of block codes employ finite Galois field (GF) algebraic concepts.

There are many different families of block codes. Because of their ease of code synthesis and implementation, for ATC VHF A/G systems application, we restrict our attention to the family of linear block codes. A linear code distinguishes itself from a nonlinear code by the property that any two codewords of a linear code can be added in "modulo" arithmetic to produce a third codeword in the code. The codes used in practical applications are almost always linear codes. This memorandum will be concentrated on some important classes of block codes including Golay codes, Bose-Chaudhuri-

Hocquenghem (BCH) codes and Reed-Solomon (RS) codes. Specifically, half-rate (or approximately 1/2) codes including the extended Golay(24,12), Golay(23,12), BCH(31,16) and BCH(127,64) codes , and high rate codes including the BCH(2040,1992), BCH(2295,1992) and RS(255,249) codes are examined.

Golay Codes

The Golay code is a binary linear (23,12) code that is capable of correcting any combination of three or fewer random errors in a block of 23 bits. The code has a minimum distance of 7. The distance (Hamming) between two codewords is the number of corresponding elements or positions in which they differ. The smallest value among the pairwise distances between all pairs of codewords is called the minimum distance of the code. The extended Golay code is obtained by adding an overall parity to the (23,12) code to become a binary (24,12) code. The added parity bit increases the minimum distance d_{min} from 7 to 8 and produces a rate 1/2 code, which is easier to implement (with regard to system clocks) than the rate 12/23 Golay code. The extended Golay (24,12) code can correct all triple errors and some (but not all) four-error patterns. It is one of the more useful and widely used linear block codes.

Bose-Chaudhuri-Hocquenghem (BCH) Codes

BCH codes are a powerful class of cyclic codes that provide a large selection of block lengths, code rates, alphabet sizes, and error-correcting capabilities. The most commonly used BCH codes are characterized as follows. Specifically, for any positive integers m (equal to or greater than 3) and t [less than $(2^m - 1)$], there exists a binary BCH(n,k) code with the following parameters:

Block length:	$n = 2^m - 1$
Number of message bits	$k \geq n - m\,t$
Parity check size	$n - k$
Minimum distance:	$d_{min} \geq 2t + 1$

Each BCH code is a t-error correcting code in that it can detect and correct up to t random errors per codeword. The BCH codes offer flexibility in the choice of code

parameters, namely, the block length and the code rate. In our examples, BCH(31,16) corrects up to 3 random errors while BCH(127,64) corrects up to 10 random errors. Both are approximately 1/2 rate codes. For a fixed m the higher the number of the redundant bits, (n - k), the more errors the BCH code can correct and the higher the resulting bandwidth expansion.

Reed-Solomon Code

The RS codes are nonbinary BCH codes. The encoder for an RS code differs from a binary encoder in that it operates on multiple bits rather than individual bits. Specifically, the encoder for an RS(n,k) code uses m-bit symbols and groups the incoming binary data stream into blocks, each k•m bits long. Each block is treated as k symbols, with each symbol having m bits. The encoding algorithm expands a block of k symbols to n symbols by adding (n - k) redundant m-bit symbols. A popular value of m is 8. 8-bit symbols are called bytes or octets. A t-symbol-correcting RS code of m-bit symbols has following parameters:

Block size:	$n = (2^m - 1)$ symbols
Message size:	k symbols
Parity-check size:	$n - k = 2t$ symbols
Minimum distance:	$d_{min} = (2t + 1)$ symbols

The RS(n,k) code can correct up to t symbols per codeword. RS codes achieve the largest possible minimum distance for any linear code with the same values of k and n. Another reason for the importance of the RS codes is the existence of an efficient hard-decision algorithm which makes it possible to implement relatively long codes in practical applications [1].

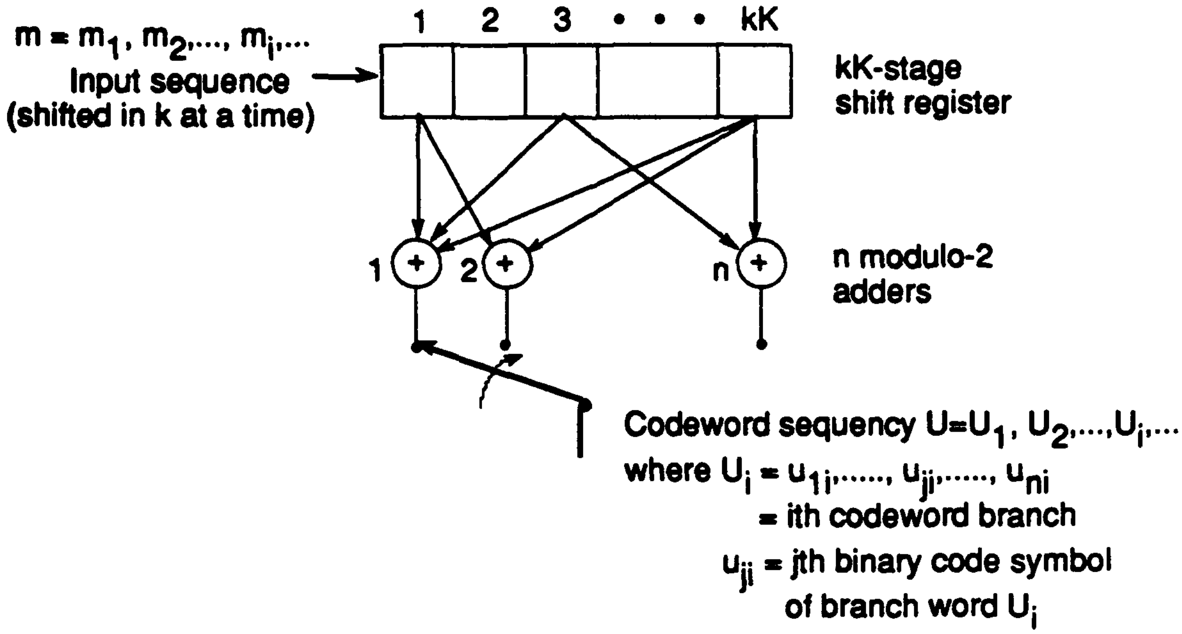
The RS(255,249) code was proposed by ARINC for a VHF digital radio [2]. The RS(255,249) code is a 256-ary linear block code utilizing 8-bit symbols. It has a minimum distance equal to 7 and can correct up to 3 codeword symbol errors. In the RS (255,249) coding, transmissions are split into 249 octets (1992 bits). The six parity check octets are then appended to the 249 octets of the data to produce a 255 octet block (2040 bits).

A 256-ary code is particularly matched (in the sense of obtaining the maximum coding gain) to a 256-ary modulation technique for transmitting the 256 possible symbols. The optimum demodulator for such a signal consists of 256 matched filters (or cross correlators). Since the candidate digital modulation scheme for the ATC A/G VHF systems in the near-term appears to be OQPSI modulation or its derivatives which basically are binary channels, a tradeoff assessment between the 256-ary RS code and its corresponding binary code on the E_b/N_0 performance under the binary channel is necessary. The corresponding binary codes of a 2^m-ary linear code can be constructed by mapping the linear codes in $GF(2^m)$ into linear codes in GF(2). An (n, k), $d_{min} = 7$ RS code over $GF(2^8)$ can be mapped into a (mn, mk), $d_{min} = 7$ binary code or a (n+mn, mk), $d_{min} = 14$ binary code [3]. As a result, the corresponding binary (2040,1992) code will have a distance 7 correcting up to 3 random errors, and the corresponding binary (2295,1992) code will have a distance 14 correcting up to 7 random errors. The BER performance and the implementation complexity will be discussed in the tradeoff section.

Convolutional Codes

In convolutional coding, the encoder continuously operates in a serial manner on the incoming message sequence. A convolutional code is described by three integers, n, k, and K. The ratio k/n has the same code rate significance (information bit per coded bit) that it has for block codes; however, n does not define a block or codeword length as it does for block codes. The integer K is a parameter known as the constraint length; it represents the number of k-tuple stages in the encoding shift register. Figure 9-1a shows an encoder of a binary convolutional code with rate k/n, measured in information bits per symbol (or coded bit) and constraint length K. It may be viewed as a finite-state machine that consists of an K-stage shift register with prescribed connections to n modulo-2 adders, an input sequence shifted in k bits at a time and a multiplexer that serialized the outputs of the adders. Figure 9-1b shows a flow chart for an encoder for the (2,1) K=7 convolutional code frequently found in commercial satellite applications, for example, the INMARSAT aeronautical mobile system.

Decoding the FEC codes is normally more complex than encoding. For the convolutional code, decoding is most often done by the Viterbi Algorithm, although other



Source: Reference 5, p312

Figure 9-1a. Convolutional Encoder with Constraint Length K and Rate k/n

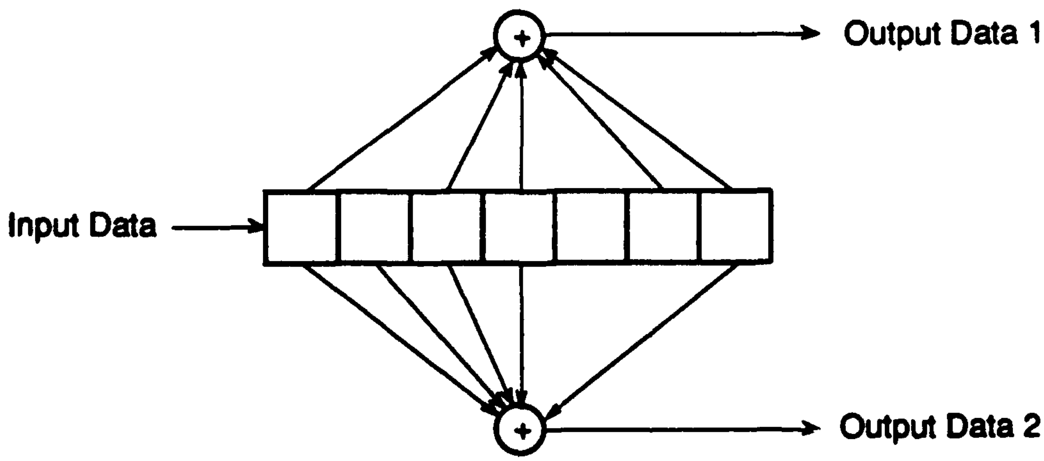


Figure 9-1b. A Rate 1/2, Constraint Length 7 Convolutional Encoder

methods are available. Whichever method is used, the decoder needs to know the history of the decoded stream (the values held in the shift registers) before being able to decode a particular bit. Therefore, with convolutional codes, if the decoder loses or make a error in the history of the stream, errors will propagate. Such situation can occur in the fading or burst interference channel. In addition, the decoder has to perform a larger number of operations per decoded bit.

Automatic Retransmission Request

Three basic types of ARQ schemes, Stop-and-Wait ARQ, Go-back-N ARQ, and Select-Repeat ARQ and a hybrid ARQ/FEC should be considered for ATC VHF A/G systems. Again, in this section, the four types of ARQ schemes are reviewed and the effects of these schemes on the issues listed earlier will be discussed in the next section.

Stop-and-Wait ARQ

With Stop-and-Wait ARQ, the transmitter sends a coded block to the receiver and waits for a positive (ACK) or negative (NAK) acknowledgment from the receiver. If an ACK is received (no error detected), the transmitter sends the next coded block. If a NAK is received (errors detected), the transmitter resends the preceding coded block. A reliable feedback channel is needed for requesting message retransmission. This remains true for all four of the ARQ schemes. The feedback channel has to be practically noiseless. This is usually not a severe restriction. Since the need for information transfer over the feedback channel is at a much lower rate, a substantial amount of redundancy can be used to ensure the reliable feedback transmissions.

Go-back-N and Select-Repeat ARQ

With Go-back-N ARQ and Select-Repeat ARQ, the transmitter sends codewords to the receiver continuously and receives acknowledgments continuously. In practice, the receiver only sends a NAK to the transmitter when an error is detected. When a NAK is received, the transmitter begins a retransmission. In the Go-back-N ARQ, the transmitter backs up to the

coded block in error and resends the block plus the blocks that follow it; while in the Select-Repeat ARQ, the transmitter simply resends those coded blocks that are acknowledged negatively.

Hybrid ARQ/FEC

The most important advantage of ARQ is that the delivered data from ARQ has predictable quality; the greatest disadvantage of ARQ is that the throughput is dependent on the channel conditions. The data throughput decreases rapidly as the channel error rate increases. It could become zero if the channel error rate causes errors in every packet received. FEC behaves in a complementary fashion. Namely, the throughput is constant, with data quality depending on the channel conditions. This suggests that combinations of FEC and ARQ could mitigate the drawbacks in both ARQ and FEC and obtain both high data throughput and reliability. The FEC scheme can be incorporated with any of three basic ARQ schemes. If a linear block code (which can be used for simultaneous error detection and error correction) is used and a received coded block is detected in error, the receiver first attempts to locate and correct the error. If the number of errors (or the length of an error burst) is within the designed error-correcting capability of the code, the errors will be corrected and the decoded message will be passed to the user or saved in a buffer until it is ready to be delivered. If an uncorrectable error pattern is detected, the receiver rejects the received codeword and requests a retransmission.

If a convolutional code is used, the retransmission strategy can be as follows: Block boundary markers can be inserted in the coded data stream so that the transmitted data stream has a block structure. This will increase the redundancy somewhat. If the decoding of a "coded block" exceeds a predetermined time limit, an erasure is declared and a retransmission of the "coded block" is requested.

9.4 TRADEOFFS

Each error control technique has pros and cons for its application to a communications system. To choose an error coding technique suitable for an ATC VHF A/G communications system, tradeoffs between various coding techniques have to be made. In

this section we examine and compare, based on the issues listed, the various error control techniques that have been introduced in the previous section.

Error Performance Improvement

The FEC coding gain is defined as the difference in the information bit energy-to-noise power spectral density ratio, E_b/N_0, required for coded and uncoded systems to provide a specified information BER when operating with the same channel conditions. In addition to the code itself, the coding gain also depends on the channel BER and the kind of modulation and demodulation scheme employed. Digital modulations of OQPSK and its derivatives, such as 4-OQAM and A-QPSK appear to be near-term candidates for the ATC VHF A/G channel (see section 8). A-QPSK is a root-raised cosine OQPSK signal with $\alpha = 1$ received into the same shaping filter, where α is the rolloff factor of the raised cosine filter. As a result, an E_b/N_0 loss is estimated at approximately 1 dB with respect to the OQPSK signal. 4-OQAM is currently a variant of OQPSK with a raised cosine, $\alpha = 0.6$, shaping filter received into a 6-pole Butterworth filter with a 3 dB bandwidth of 12 kHz. The resulting E_b/N_0 loss is believed to be in the same order as the loss in A-QPSK. To show typical error performance for various error control codes, OQPSK modulation is used in the discussion. Figure 9-2 and figure 9-3 illustrate the probability of an information bit error, P_{be}, versus E_b/N_0 for coherent OQPSK modulation with hard decisions in additive white Gaussian noise (AWGN) in combination with no code and various block coded systems discussed in this paper.

The error performance of various FEC codes was calculated as follows. Since E_b/N_0 is the uncoded data signal-to-noise ratio, the signal-to-noise ratio for coded data, E_c/N_0, is given by

$$E_c/N_0 = (k/n)\,(E_b/N_0) \qquad (1)$$

where k/n is the code rate, equal to 1/2 in the Golay (12,24) code and 249/255 in the Reed-Solomon (255,249) code, for example.

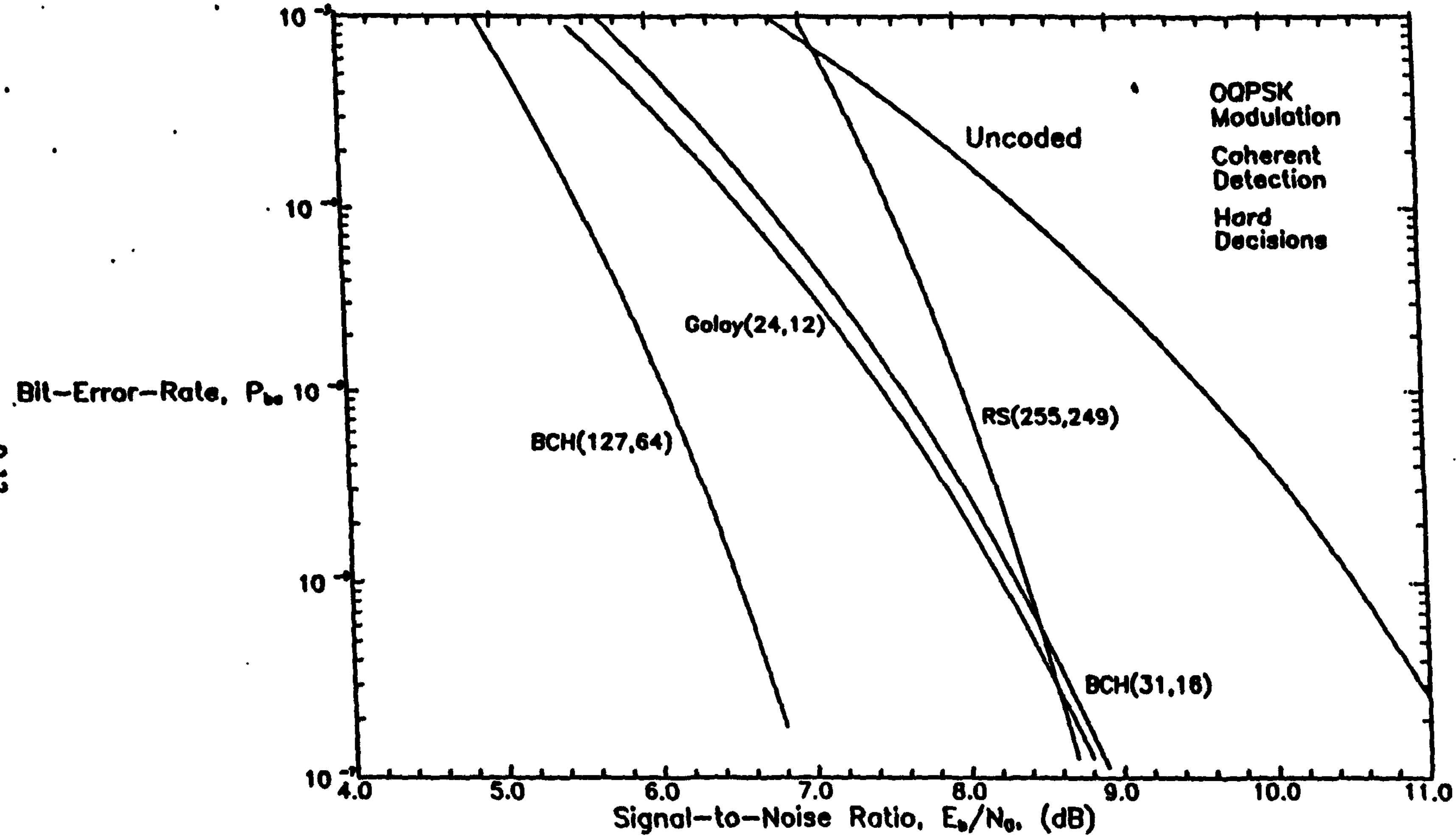


Figure 9-2. P_{be} versus E_b/N_0 over a Gaussian Channel for Several Half-Rate Binary and High-Rate RS Block Codes

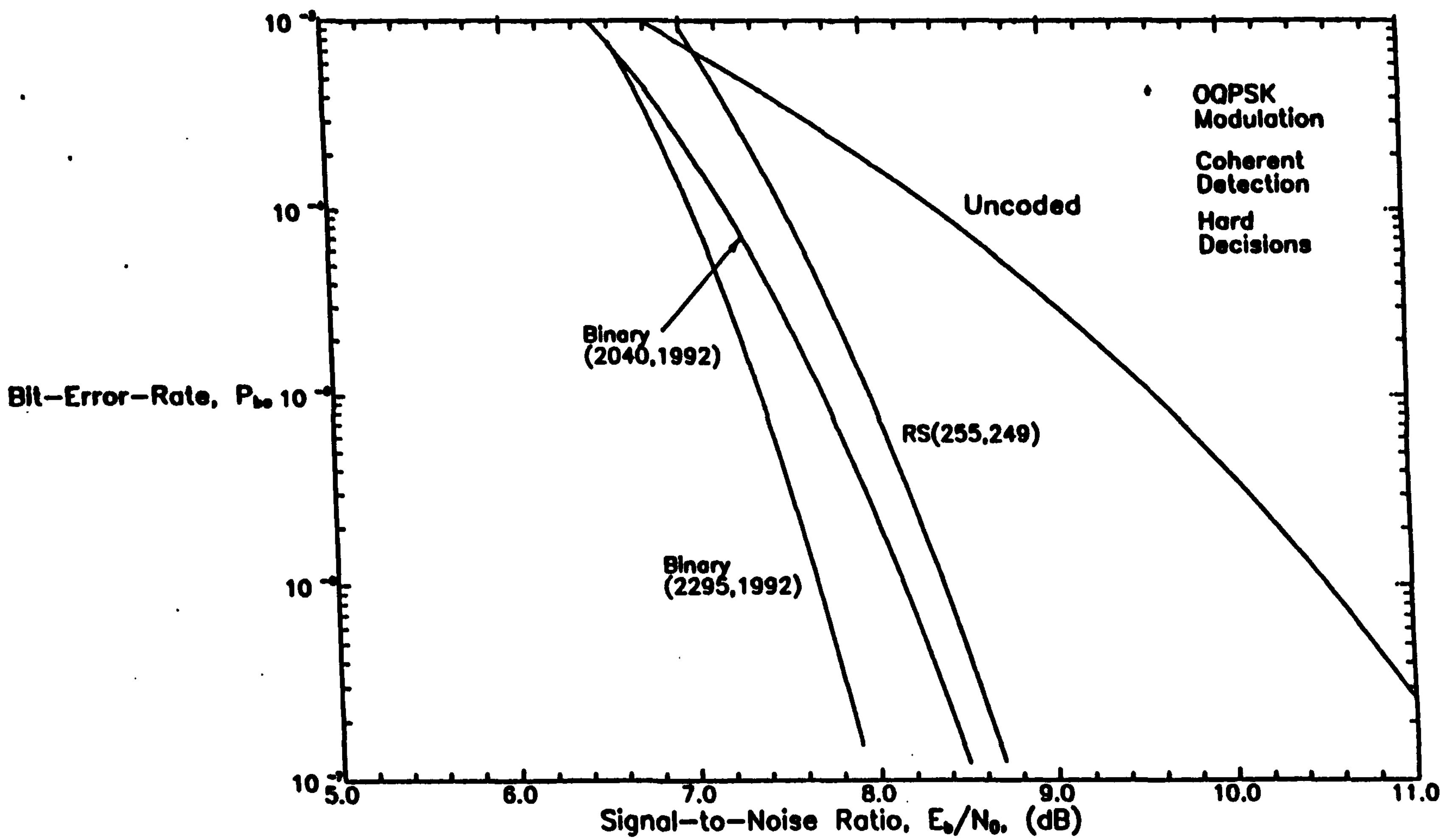


Figure 9-3. P_{be} versus E_b/N_0 over a Gaussian Channel for Several High-Rate Block Codes

OQPSK is used to modulate the coded digital data on the RF frequencies. The coded (channel) BER can be found to be [4].

$$P_c = 1/2 \ \mathrm{erfc}(E_c/N_0) \tag{2}$$

where erfc(.) is the complementary error function and is defined as

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} \exp(-y^2)\,dy .$$

The FEC decoder performance is a function of the channel symbol error probability, P_s. The decoded word symbol error probability (hard decisions), P_{sd}, is given by [5].

$$P_{sd} \equiv \sum_{i=t+1}^{n} \frac{i}{n} C_i^n P_s^i (1 - P_s)^{n-i} \tag{3}$$

where n is the length of the codeword, t is the number of the errors the code can correct and $C_i^n = \frac{n!}{(n-i)!\,i!}$. The uncoded information bit error rate, P_{be}, is related to the coded word symbol error by [6]

$$P_{be} = [2^{m-1}/(2^m - 1)]\, P_{sd} . \tag{4}$$

In a binary code, P_s equals P_c while in the 2^8-ary RS(249,255) case, a symbol has 8 bits. Assuming independent bit errors and hard decisions, the probability of a 8-bit coded symbol error is then

$$P_s = 1 - (1 - P_c)^8 . \tag{5}$$

The coding performance, P_{be} versus E_b/N_0, of the coded case is calculated from Equations (1) through (5). For the uncoded case, P_{be} versus E_b/N_0 can be obtained by Equation (2) due to the fact that $P_{be} = P_c$ and $E_b/N_0 = E_c/N_0$.

There are two types of decoding, hard-decision decoding and soft-decision decoding. In hard-decision decoding, the demodulator output consists of the discrete symbol elements. The demodulator feeds the symbols to the decoder which operates on the hard decisions made by the demodulator. In the soft decision decoding, the demodulator output consists of a real number or its quantized approximation (with greater than two quantization level). The demodulator feeds such quantized code symbols to the decoder and the decoder operates on the soft decisions made by the demodulator. In the case of hard-decision decoding, a single bit is used to describe each decoder input sample, while for eight-level quantized soft-decision decoding, for example, 3 bits are used to describe each decoder input sample; therefore, three times the amount of data must be handled or stored during the decoding process. Hence the price paid for soft-decision decoding is an increase in required speed and memory size at the decoder. As a result, block codes (especially with long codewords, i.e., large n) with soft-decision decoders are substantially more complex than hard-decision decoders; therefore, block codes are usually implemented with hard-decision decoders.

For convolutional codes, both hard- and soft-decision implementations are equally popular. The most prevalent use of the soft-decision decoding is with the Viterbi convolutional decoding algorithm since with Viterbi decoding, soft decisions represent only a trivial increase in computation. At a typical BER of 10^{-5} for data applications, soft-decision decoding will result in a coding gain approximately 2 dB larger than for hard-decision decoding.

Binary FEC codes obtain their optimal BER performance in combination with binary digital modulations such as BPSK. The RS(255,249) code is a 2^8-ary symbol linear block code. It has a minimum distance, $d_{min}=7$. It can correct up to 3 coded-word symbol errors. Its optimal coding performance is achieved by modulating the data with a matched 2^8-ary signaling channel. For example, the OQPSK modulation technique has 4 channel symbols and therefore, the coding performance of the RS(255,249) code will not be optimal due to the unmatched modulation scheme employed. Figure 9-3 shows the performance of RS(255,249) and its corresponding binary (2040,1992) and (2295,1992) codes with OQPSK modulation. The two binary codes are more nearly matched to OQPSK quaternary modulation. The error performance for all the block codes shown in figures 2 and 3 are with hard decisions. The coding gains of block codes for BERs equal to 10^{-4}, 10^{-5}, and 10^{-6} are shown in Table 9-1.

From Table 9-1, we see that RS(255,249) and its corresponding binary codes have about the same error performance at BER=10^{-5} as the extended Golay(24,12) code or BCH(31,16) code. The more complex implementation of the longer codes does achieve better bandwidth efficiency reduction (e.g., rate 249/255 for the RS code versus rate 1/2 for the shorter codes). The penalty in coding performance of RS(255,249) code due to the unmatched modulation scheme is insignificant, approximately 0.3 dB compared to the binary(2040,1992) code and 0.5 dB compared to the binary(2295,1992) code. For BCH codes having a code rate of approximately 1/2, the code with a longer codeword (more redundant bits) has a better coding gain.

Table 9-2 lists the coding gains for a few popular commonly used convolutional codes employing Viterbi decoding. Viterbi decoding is normally used in decoding the convolutional codes with hard- and soft-decision decoding equally popular. One advantage of the Viterbi-decoded convolutional coding systems is that they can easily take advantage of soft decision data. From both Table 9-1 and Table 9-2, we see that the coding gains increase as the BER is decreased. With the same code rate and in the ATC required BER range (10^{-2} to 10^{-6}), the convolutional code with soft decisions outperforms the block codes in the AWGN channel.

The data reliability delivered by an ARQ scheme depends on its error detection capability. The error detection capability of ARQ depends on the length of cyclic redundancy check (CRC) code employed. In both ATS and AOC, both the airborne and ground radios are expected to be capable of interfacing to other equipment or sub-networks comprising the Aeronautical Telecommunication Network (ATN). In the packet data communication mode, the current MASPS draft document requires a maximum undetected packet error rate of 10^{-6} for a packet 128 octets (1024 bits) in length. If ARQ with a CRC code is not used, this implies that a BER of about 10^{-9} is required, assuming the bit errors occur randomly. This BER is too stringent for a practical digital modulation scheme and FEC code combination, especially over a typical ATC VHF A/G channel. For transmissions coded with a CRC code for error detection and an ARQ mechanism for data retransmission,

Table 9-1. Bandwidth Efficiency and Power Efficiency of Block Codes with OQPSK Modulation

Coding Schemes	Bandwidth Efficiency (b/s/Hz)	Power Efficiency Eb/No (dB) for Bit-Error-Rate 10^{-4} (Coding Gain)	10^{-5} (Coding Gain)	10^{-6} (Coding Gain)
Uncoded	2.00	8.4	9.6	10.5
Golay (24,12) Hard Decisions	1.00	6.6 (1.8)	7.5 (2.1)	8.2 (2.3)
BCH (31,16) Hard Decisions	1.03	5.8 (1.6)	7.6 (2.0)	8.3 (2.2)
BCH (127,64) Hard Decisions	1.00	5.6 (2.8)	6.3 (3.3)	6.5 (4.0)
RS (255,249) Hard Decisions	1.95	7.53 (0.87)	8.0 (1.6)	8.35 (2.15)
BCH (2040,1992)* Hard Decisions	1.95	7.2 (1.2)	7.75 (1.85)	8.2 (2.4)
BCH (2295,1992)** Hard Decisions	1.73	7.0 (1.4)	7.5 (2.1)	7.6 (2.9)

* 256-ary Reed Solomon (255,249) code maps to the binary code without distance change.
** 256-ary Reed Solomon (255,249) code maps to the binary code with distance doubled.

Table 9-2. Bandwidth Efficiency and Coding Gain for Convolutional Codes with OQPSK Modulation and Viterbi Decoding

Coding Schemes	Bandwidth Efficiency (b/s/Hz)	Coding Gain (dB) for Bit-Error-Rate		
		10^{-4}	10^{-5}	10^{-6}
Convolutional Rate 1/2, K=7 Hard Decisions	1.00	2.7	3.1	3.4
Convolutional Rate 1/2, K=7 3-bit Soft Decisions	1.00	4.65	5.2	5.6
Convolutional Rate 3/4, K=9 Hard Decisions	1.50	1.9	2.4	2.6
Convolutional Rate 3/4, K=9 3-bit Soft Decisions	1.50	3.7	4.25	4.6

the error performance has been evaluated [7] and is shown in figure 9-4. To achieve a undetected packet error rate of 10^{-6}, a 3×10^{-5} BER is required, which can be obtained by a practical digital modulation combined with a FEC code. The main concern of an ARQ scheme is its data throughput, and this is discussed in the next section.

Bandwidth Efficiency and Data Throughput

To generate an (n,k) block code, the encoder accepts information in successive k-bit blocks; for each block, it adds (n-k) redundant bits that are algebraically related to the k message bits, thereby producing an overall encoded block of length n bits. The code rate, k/n can be thought of as the portion of a coded bit that constitutes information. The encoder produces bits at the rate $R_c = (n/k) R$, where R is the information rate generated by the source. The bit rate R_c is the channel data rate and therefore, the rate n/k represents the channel bandwidth expansion. Thus the bandwidth efficiency in b/s/Hz, defined as the number of information bits transmitted per channel symbol of a (n,k) block code, is reduced by the factor k/n. For example, an error control technique that employs rate 1/2 code (100% redundancy) requires twice the bandwidth of an uncoded system. However, if a rate 3/4 code is used, the redundancy is 33% and the bandwidth expansion factor is 4/3.

The bandwidth efficiencies of various block and convolutional codes are given in Table 9-1 and Table 9-2, respectively. For a fixed code length, n, a bandwidth efficient code has a large k value which, in turn, implies that the code has fewer redundant bits and a reduced error correcting capability. Therefore, high bandwidth efficient codes having good error correcting capabilities usually have large n and k values. However, the computational requirement and the implementation complexity of the Viterbi algorithm of convolutional decoding restricts its practical applications to codes with small constraint length, such as 7 or 9, and rate 3/4 and 1/2 or lower. This is because to increase the code rate of a convolutional code requires increasing the number of state variables which could exponentially increase the implementation complexity of a Viterbi decoder.

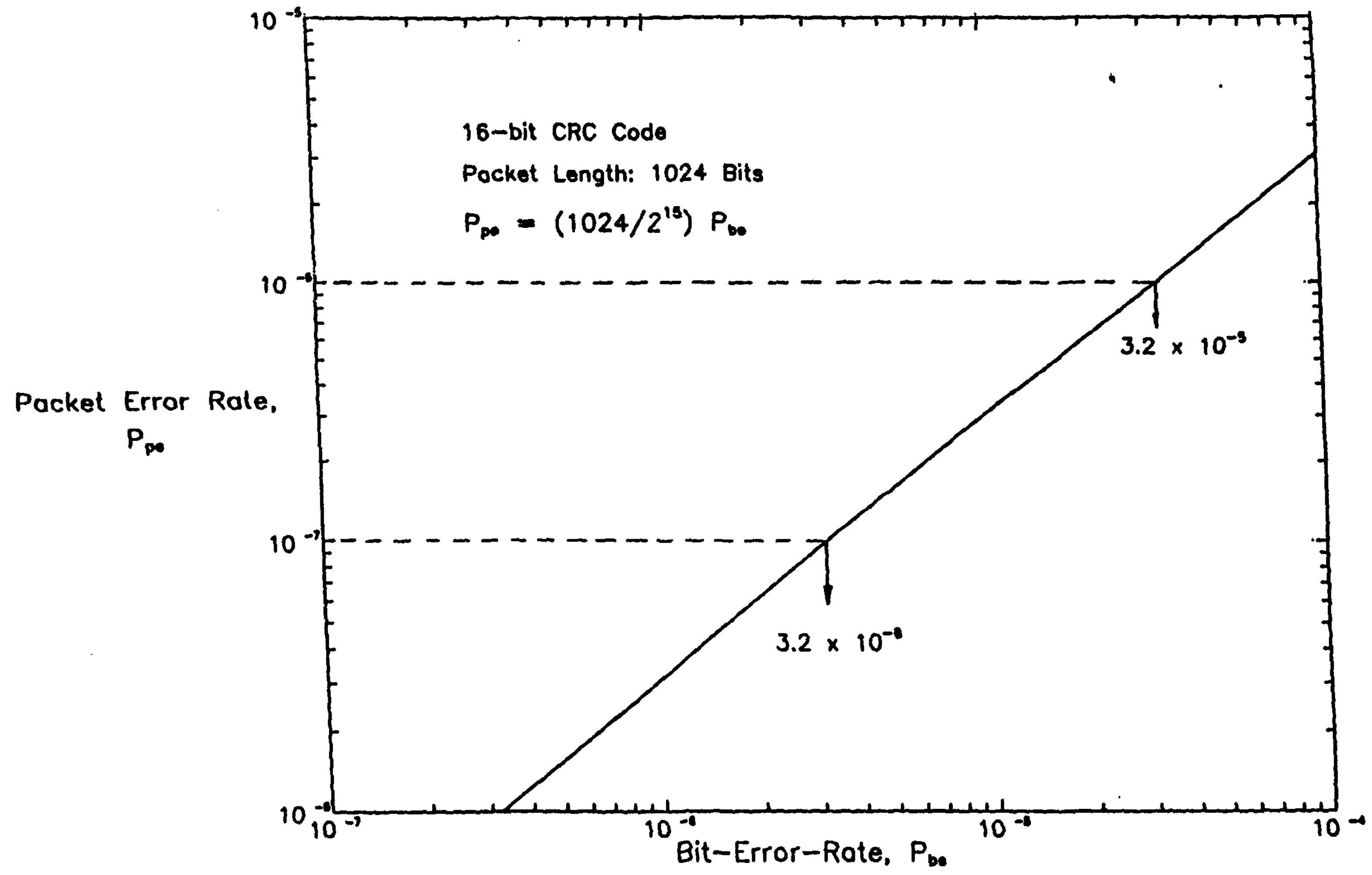


Figure 9-4. Packet (Undetected) Error Rate Versus Channel Bit-Error-Rate

The performance of an ARQ error control scheme is normally measured by its reliability and throughput efficiency. The reliability of an ARQ is determined by the undetected packet error rate which translates into a corresponding channel BER depending on the packet length and the ARQ strategy. The throughput efficiency is defined as the ratio of the average number of information bits per second accepted at the receiver to the maximum data transmission rate on the channel.

An obvious problem with Stop-and-Wait ARQ is that while the transmitter is idling, waiting for an acknowledgment, transmission time is wasted and throughput suffers, especially when the round-trip delay is long. The round-trip delay is defined as the time interval between the start of transmission of a packet and the decoding of an acknowledgement for that packet.

In a Go-back-N system, the receiver requests that all of the last N blocks are repeated whenever a block is detected in error. The advantage of a Go-back-N system is that the blocks do not have to be individually labeled and that the algorithms are accordingly simpler than a comparable Select-Repeat scheme. In fact, the lack of numbering leads to less overhead and allows shorter blocks which, in turn, means that the block error performance in a noisy environment on the feedback link can be slightly better. However, there are disadvantages. First, many of the blocks that are retransmitted may have already been successfully received, i.e., as many as all (N-1) blocks following the one that was received with a detected error. Secondly, in order to ensure that erroneous blocks are repeated, the maximum round-trip delay on the transmit-receive path is constrained. The delay can not exceed the time it takes to transmits N blocks. This is one consideration for the ATC VHF A/G communications application.

In a Select-Repeat system, the receiver asks for specific blocks to be repeated. The advantage of a Select-Repeat system is that only blocks with errors are repeated. In a Select-Repeat system, blocks have to be individually labeled. This means that the overhead is either higher than a comparable Go-back-N system or that the blocks are longer. Longer blocks are likely to contain more errors than a shorter block for the same channel conditions. This presents a tradeoff between minimizing the loss of throughput due to overhead and minimizing the loss of throughput due to repeats. Select-Repeat systems are more complex than Go-back-N systems but they are much less restricted with respect to the round-trip

delay. However, some limitations still exist. The blocks have to be numbered individually and the length of the block counter upper bounds the maximum round-trip delay. Another limitation is on the amount of memory that the transmitter has available to store transmitted blocks.

The throughput of the three basic ARQ schemes has been analyzed in detail [8]. The throughput depends on the transmission rate, the packet size, the channel characteristics, the size of the data buffers, the round-trip time delay, and the protocol rules of the particular scheme. The analysis indicates that the Select-Repeat ARQ is the most efficient, whereas the Stop-and-Wait ARQ is the least efficient. The throughput of the Select-Repeat ARQ (with infinite data buffer) does not depend strongly on the round-trip delay of the system; however, the throughputs of the other two schemes depend on the round-trip delay. In systems where the round-trip delay is large and data rate is high (mega bits per second), the throughput for the Go-back-N ARQ drops rapidly as the channel error rate increases, while the throughput of the Stop-and-Wait ARQ becomes unacceptable.

The high throughput of the Select-Repeat ARQ is achieved at the expense of more overhead, extensive buffering at the receiver and more complex logic at both the receiver and transmitter. Limitation of the buffer length in the implementation might reduce the throughput of the system. However, in general, the Select-Repeat ARQ outperforms the other two schemes in systems where data transmission is high and round-trip delay is large.

In the ATC situation, the maximum information rate is expected to be around 40 kb/s (which is a moderate transmission rate) and the maximum packet duration could take as long as 500 ms (by assuming a maximum packet length of approximately 10,000 bits and a minimum rate of about 20 kb/s). The maximum round-trip propagation delay is approximately 4.2 ms. These parameters are derived in section 11. The delay requirement for a system can serve as an important discriminator among ARQ schemes. For example, in real-time voice, an overall delay of no more than 100 ms may be acceptable because such a short delay is almost imperceptible to a human listener. For a packet length of 1024 bits at 42 kb/s transmission rate, the packet duration is approximately 24 ms which is much longer than the propagation delay and constitutes most of the round-trip delay. Therefore, a possible value of N for Go-back-N ARQ is 2. The throughput efficiencies of the three basic ARQ schemes under these conditions are calculated and shown in figure 9-5. From figure 9-5

we see that the Stop-and-Wait ARQ is inefficient. For a data mode at BER=3×10^{-5}, the throughput of Select-Repeat ARQ is approximately 97% and that of Go-back-N ARQ is 94%. For a voice mode at BER=10^{-3}, the throughput of Select-Repeat ARQ is approximately only 36% while that of the Go-back-N ARQ is only 22%.

All three basic ARQ systems have the same general performance, i.e., for a good channel (low channel error rate), the throughput is high, and for very poor channels, throughput is low. The use of the FEC in the hybrid ARQ/FEC scheme increases the range of channel BER for which throughput remains high, i.e., the FEC reduces the number of retransmissions that are required. A price is paid in code rate and complexity, but if the forward error correcting code significantly reduces the number of retransmissions, throughput is increased.

Code Robustness

The FEC codes discussed in the last section work well in the channels which cause independent errors in the received codeword. AWGN channels, such as satellite and deep-space communications links produce independent errors.

Multipath exists when there is more than one transmission path between transmitter and receiver. In the VHF ATC A/G radio environment, multiple random propagation paths in addition to the primary direct path may occur. Consequently, multipath fading may appear on the received signal and cause burst errors in addition to the independent errors.

2^m-ary RS codes can combat combinations of random and burst errors due to its m-bits symbol structure. Binary BCH codes only correct random independent errors. However, interleaving combined with error control coding can be used with binary BCH codes or 2^m-ary RS codes in the burst-error channel. Interleaving is a straightforward and effective method to effectively transform a bursty channel into an independent-error channel for which many FEC coding techniques are applicable. Thus, a preferred transmitter for ATC application may include an encoder followed by an interleaver that scrambles the encoded

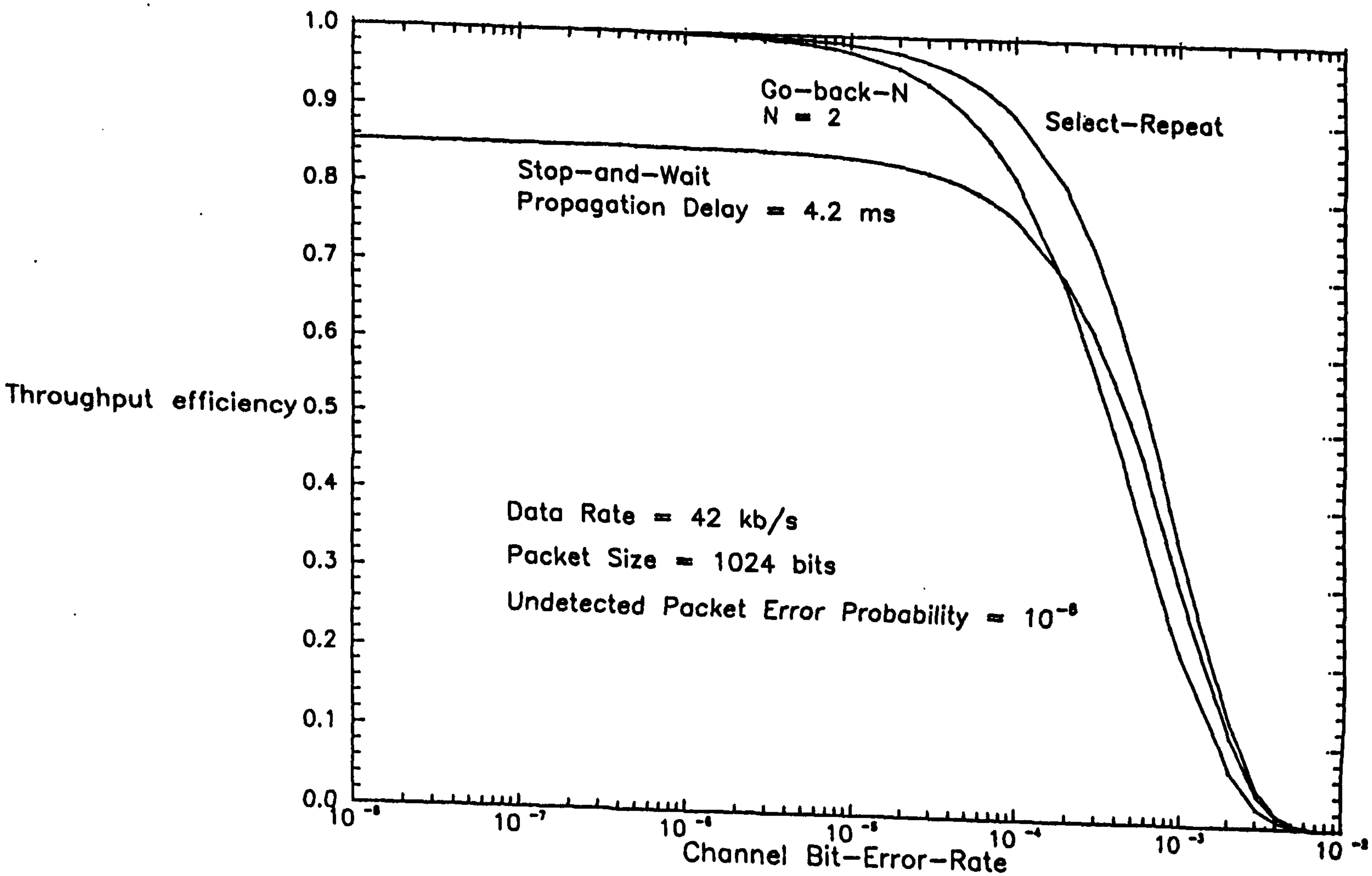


Figure 9-5. Throughput Efficiency of Three Basic ARQ Schemes

data stream in a deterministic manner. In the receiver, prior to decoding a deinterleaver is used to perform the inverse operation. Error bursts that occur on the channel are spread out in the data sequence to be decoded, causing the error burst to span many codewords. The combination of interleaving and FEC thus provides an effective means of combating the effect of error bursts. However, the data round-trip delay requirement in the ATC transmissions limits the allowable delay due to interleaving and deinterleaving processing in the transmitter and receiver and therefore, the length of the interleaving.

Table 9-3 shows the E_b/N_0 performance at BER = 10^{-6} for a rate 1/2, constraint length 7, Viterbi hard-decision decoded convolutional code in the Rayleigh multipath fading channel. Theoretically, multipath fading can be modeled by assuming Rayleigh fading statistics or Rician fading statistics. In Rayleigh fading, there are a large number of signals which vary independently, with amplitudes roughly equal and phases approximately random; while in Rician fading statistics, in addition to the desired signal from the direct path, there is a strong interference arising from another relatively stable path. In ATC A/G communication channels, the secondary transmission paths from buildings, earth reflection, foliage, etc., can be characterized by the summation of several delayed and attenuated replicas of the desired signal. Thus, the channel fading statistics may lean toward the Rician fading channel instead of the Rayleigh fading channel.

Although the ATC multipath channel is not as severe as the Rayleigh fading channel, Table 9-3 is intended to show the E_b/N_0 improvement that can be achieved by a typical FEC code in a multipath fading channel. From Table 9-3, we see that a large coding gain, approximately 37 dB is achieved by applying an FEC code in combination with OQPSK modulation to the Rayleigh fading channel.

ARQ is an effective method for error protection over the burst error channels. The error detection capability of ARQ increases with the length of CRC code employed. A CRC code with a length b will detect all burst errors of length b or less. But a longer CRC code increases the redundant bits in the data stream and therefore, reduces the b/s/Hz bandwidth efficiency. ARQ should be considered for the multipath fading channel. The length of the CRC code to be employed depends on the undetected packet error rate required.

Table 9-3. Bit-Error-Rate Performance of Rate 1/2, K=7, Hard Decision Convolutional Code with Viterbi Decoding in Rayleigh Fading

Coding Scheme	Bandwidth Efficiency	Power Efficiency Eb/No for BER = 10^{-6} in dB
Uncoded OQPSK Gaussian Channel	2.00	10.5
Coded OQPSK Gaussian Channel	1.00	7.1
Uncoded OQPSK Rayleigh Channel	2.00	53.0
Coded OQPSK Rayleigh Channel	1.00	16.4

Implementation complexity

In FEC codes, the implementation of the decoder is usually more complex than the encoder. In general, the implementation complexity depends on the code length, number of redundant bits, and the decoding algorithm employed, particularly the number of errors that can be corrected within each codeword.

A block code with a long codeword and a high code rate requires larger memory buffers, more complex computations and longer time to encode and decode when compared to a block code with shorter codeword. The computation and memory requirement with soft decisions is further expanded and is proportional to the number of levels (number of bits) used. For example, computation time and memory size could be tripled or more with 3-bit soft decisions when compared to hard decisions. Convolutional decoders have a high complexity in terms of decoding operations per output bit. Furthermore, a majority of these operations involve addition and comparison of real numbers. The main factors governing the implementation complexity of a Viterbi decoder are the constraint length, the code rate and the speed. The Viterbi algorithm is only economical for small constraint lengths such as K=7 or 9. By contrast, block codes have a smaller number of decoding operations per output bit. Moreover, RS decoders normally operate directly on symbols and real number arithmetic is avoided. When soft decisions are available the implementation complexity of a convolutional coding system is, in general, less than that of a block coding system that achieves the same BER performance.

ARQ error detection capability in practice is implemented with CRC codes. A CRC code is a cyclic code used for error detection. It is relatively easy to encode and decode using modern integrated circuit techniques. The implementation complexity of ARQ mainly comes from the protocol employed. ARQ with Select-Repeat is the most complex of the three basic ARQ strategies, requiring the most involved controllers and data buffers and providing the best performance. However, it is not constrained by the data rate and the round-trip delay. It is a plus in the ATC applications.

9.5 IMPACT/IMPORTANCE

Error control techniques can offer significant performance gains in a system where other components would be expensive to upgrade. An example is the applications of FEC to the satellite communications and deep space probe communications in the past decades.

In the digital modulation decision tree paper, it has been pointed out that, for a particular modulation scheme, the BER decreases as E_b/N_0 increases. Practical considerations usually place a limit on the value that we can assign to E_b/N_0. We often arrive at a modulation scheme and find that under certain channel situations, the E_b/N_0 is not high enough for the modulation to provide the needed data quality. Since error control coding can provide improved data performance at a given E_b/N_0, it is one practical option available for changing data quality from problematic to acceptable. Furthermore, if the ATC VHF A/G communications has no problem with E_b/N_0 value, the error control coding can be used to replace the expensive components in the system design to allow larger component losses while overcoming the resulting loss in performance.

9.6 TRANSITION

The current AM radio has no channel coding or error control techniques implemented for either voice or data (but the ACARS data modem does). The voice modulation is analog DSBTC and the data transmission is accomplished by using an MSK signal to AM modulate the voice radio. The future improved ATC VHF radio may well include digital voice and digital data modulation.

The new system should provide for coexistence with the current system and a gradual phasing in of the new system. Since error control coding is not particularly effective at BERs that can be tolerated by many digital voice techniques, the ATC VHF voice channel may not use error control coding. However, error control techniques would apply to the data link. Since an error control coding technique would need to be combined with the digital modulation to provide more reliable data link communications and alternate routing and networking, the error control technique would be introduced concurrently with the digital modulation technique.

9.7 CRITERIA FOR DECISION

There are a multitude of error control coding techniques which could be considered for ATC VHF A/G systems. The final choice will depend on the priority given to each of the following goals: 1) to obtain a maximum coding gain so that any excess E_b/N_0 can be used to minimize the required transmitter power on a given link; 2) to minimize the reduction in bandwidth efficiency and information throughput; 3) to maximize the performance improvement in the slow and fast multipath channels; and 4) to minimize the design complexity.

In light of these goals and the issues and tradeoffs discussed earlier, some general conclusions can be made as follows:

1) ATC channels can exhibit a mixture of independent and burst errors. Burst of errors are produced by atmospherics, multipath fading, and interferences from other users of the frequency band. Either ARQ or FEC alone may not be sufficient to provide a cost-effective reliable ATC VHF A/G data communications. A hybrid ARQ and FEC strategy with data interleaving may very likely be the choice.

2) The data throughput of the Select-Repeat ARQ is the most efficient among the three basic ARQ schemes, and its performance is the least constrained by the data round-trip-delay. In general, the Select-Repeat ARQ outperforms the other two schemes in systems where the data rate and round-trip delay are large. In the ATC situation, the transmission rate is moderate. For data mode application, the performance of Select-Repeat ARQ is slightly better than the Go-back-N ARQ. Since more overhead, and more logic at both receiver and transmitter are needed in the Select-Repeat ARQ, the choice can be either Select-Repeat ARQ or Go-back-N ARQ. However, for voice modes, the Select-Repeat ARQ performs much better than Go-back-N. The preferred choice appears to be the Select-Repeat ARQ scheme.

3) Convolutional coding is a preferred method in situations where the message bits arrive serially rather than in large blocks when the use of a buffer may be undesirable. In block coding, the encoder accepts a k-bit message block and generates an n-bit codeword. The codewords are produced on a block-by-block basis. In situations

where messages are treated in packets, block codes may be preferred for both error detection and error correction due to their blocked data format. In addition, if the ATC data transmission only allows very little redundancy for the error control coding, a high rate linear block code such as RS(255,249) is preferred over the convolutional code, since a convolutional code with Viterbi decoding is only economical for small constraint lengths such as 7 or 9 which confines the code rate to 3/4 and 1/2 or lower. In general, the convolutional code complexity increases as the redundancy decreases; while for block codes the complexity decreases as the redundancy decreases.

4) From the ability in combating channel errors viewpoint, both convolutional codes and binary BCH codes are weak when it comes to burst errors, and RS codes are superior. RS codes can combat combinations of both random and burst errors. Convolutional codes and binary BCH codes are good when the noise is white and Gaussian. Although the data interleaving technique can be used to transform the burst errors into random independent errors, the length of data interleaving needed in deep fade situations might cause excessive delay due to the interleaving and deinterleaving processing. To combat burst errors which are expected in the ATC channel by using convolutional codes or binary BCH codes with data interleaving may not be a good choice. In the ATC application, the family of RS codes with the data interleaving appears to be a better choice.

5) RS codes are symbol based codes rather than binary codes so that the encoder and decoder logic work with byte-based arithmetic. This reduces the complexity of the logic as compared to a BCH binary code of the same length.

9.8 CONNECTIVITY/RELATIONSHIP WITH OTHER DECISIONS

Section 9 has discussed the choices among error control techniques for VHF A/G systems, branchpoint 4.1.2.2 on the decision tree. Section 8 has discussed the choices among the digital modulations, branchpoint 4.1.2 1 on the decision tree. Both topics form the branch 4.1.2 labeled Digital Modulation/Coding. In a sense, error control techniques deal with transforming digital data streams into "better sequences" so that errors can be detected and removed after the signals are demodulated. Digital modulation techniques deal with transforming waveforms into "better waveforms" to make the detection process less subject to error. The goals behind both techniques are to improve E_b/N_0 performance and to provide efficient and reliable data communications.

SECTION 10

MULTIPLEXING

10.1 CONTEXT

A systematic process for examining technical alternatives for improved air/ground communications in air traffic management has been established. A decision tree structure is shown in Appendix A that attempts to organize various alternatives in a top-down hierarchy. This provides a framework for evaluating potential solutions that can be represented by paths through the decision tree.

This paper addresses a particular alternative in the attached decision tree, namely, branchpoint 4.3.2, Multiplexing. The organization of this paper is as follows: Background, Issues, Tradeoffs, Impact/Importance, Criteria for Decision, and Connectivity/Relationship with Other Decisions. The different sections will also deal with issues related to the transition from the current system to new alternative ones where appropriate.

10.2 BACKGROUND

One of the major reasons for considering an upgrade to the VHF air/ground voice communications component of the air traffic control (ATC) system is to increase the system's capacity. In view of an anticipated increase in aircraft traffic density, there is a need to squeeze as many channels as possible into a limited spectrum allocation. In a general sense, the different types of multiplexing are just different ways of dealing with this issue: how to let many communications channels share a given transmission medium efficiently.

The current ATC system already employs a simple (but effective) form of frequency division multiplexing. Each controller within a given geographical area has exclusive rights to one or more frequencies which are well-separated from the frequencies used by other controllers in the region. Currently, the frequencies are spread across the frequency band from 118 MHz to 137 MHz on 25 kHz centers (50 kHz centers are still most common). These links use double sideband AM whose bandwidth is on the order of 7 kHz. The use of more efficient analog modulations has been discussed in section 6.

Note that the current system also reuses frequencies in well-separated geographical areas. This type of frequency reuse might be termed "space division multiplexing". It is possible that efficiency could be significantly enhanced by taking a more systematic approach to frequency reuse. This aspect of efficient spectrum usage will be covered at length in a separate paper and will not be the focus of this one.

The focus of this paper will be on techniques which manipulate communications in the frequency and time domains. In addition to the previously-mentioned frequency division multiplexing (FDM), well-known multiplexing methods include time division (TDM), and code division (CDM). These will be described in the next section along with related issues of system requirements and constraints.

10.3 ISSUES

In this section the different multiplexing techniques are explained within the context of their application to ATC air/ground voice communications. In order to clarify the discussion it would be worthwhile to define some of the terminology used. The word "multiplexing" refers to techniques which allow multiple communications circuits to share a single medium (in this case, the spectrum between 118 and 137 MHz). This is distinguished from "multiple access" which refers to the means by which multiple users share circuits. During the discussion, the terms full duplex and half duplex are used. In standard communications parlance[1, 2] duplex means the ability to transmit and receive (effectively) simultaneously. A half-duplex radio can receive and transmit, but not simultaneously. A simplex radio can perform only one function — usually receive. Finally, the term "service channel" is used in a number of places. A service channel is a special channel used to facilitate functions like network setup and automatic frequency changes. It is analogous to the "setup channel" used in the mobile telephone system [3].

Frequency Division Multiplexing

The starting point for this discussion is the current system which can be thought of as a frequency multiplexed system with frequency channels separated by 25 kHz. Since the intrinsic bandwidth of a double sideband AM system is about 7 kHz, it appears that there is a possibility to fit three channels in 25 kHz. The current system has wider channel separation

in order to account for carrier frequency errors due to poor oscillators (and to a smaller extent Doppler shifts due to aircraft motion). To allow for such frequency inaccuracies receiver bandwidths of 40 kHz (±20 kHz) are used in modern ATC radios. Double sideband AM performance is relatively immune to such frequency offsets since the audio signal-to-noise ratio in the presence of white noise is controlled primarily by the post-detection audio bandwidth and not the IF (intermediate frequency) bandwidth except in weak signal situations. Nevertheless, the width of the IF filters would make it difficult for the current radios to coexist with a new system with narrower channel spacing.

In section 6, it has been shown that other analog modulations, particularly single sideband AM and narrowband FM are viable candidates for improved voice modulation. These would allow channel spacing to be reduced by a factor of 2 to 5.

Digital techniques can also be used to reduce the band occupied by a single voice channel. The first step in digital voice transmission is the conversion of the analog voice signal into a digital bit stream. There are many algorithms which have been developed to accomplish this transformation. The ones which are most promising for the ATC application create bit streams whose bit rates are between 4.8 kbps and 16 kbps. In general, the highest-rate vocoders are most robust. For instance, a 16 kbps (Continuously Variable Slope Delta modulation, CVSD) vocoder can produce intelligible voice even with a bit error probability of 10%, while a 4.8 kbps (Codebook Excited Linear Predictive, CELP) vocoder degrades rapidly when the error rate exceeds 1%. Note that the vocoders with higher bit rates also tend to be less susceptible to acoustic noise in the environment of the speaker. The second step in digital voice transmission is radio-frequency (RF) modulation. There are many types of RF modulation which can be used. They can be compared in a number of ways including ease of implementation, noise resistance, resistance to fading, etc. For this discussion one of the most important parameters is bandwidth efficiency, which can be measured in terms of bits per second per Hertz. High quality land lines with very good signal-to-noise ratios sometimes use complicated signalling schemes to send many bits per second per Hertz. However, VHF radio links are subject to many sources of natural and man-made noise and fading, so that it is unlikely that more than 2 bits per second per Hertz will be practical. Other factors that limit bandwidth efficiency include adjacent channel interference restrictions. Typically, a certain amount of guard band must be included in a system design

to allow for such factors. Overall bandwidth efficiency may be no more than 1.6 bits per second per Hertz, i.e., 40 kbps in a 25 kHz channel.

As an example of how these factors come into play in a frequency division multiplexed system consider the case of a 6.4 kbps digital voice channel. If the bandwidth efficiency is 1.6 bps/Hz then each voice channel requires 4 kHz of bandwidth, and 6 such voice channels could be put into a single 25 kHz channel. Of course, this estimate does not take into account any allowance for frequency instabilities or Doppler shifts, or any overhead for synchronization or other channel/net maintenance functions. A more practical limit might be 4 or 5 such channels per 25 kHz.

The above discussion on FDM applies to half-duplex systems, i.e., systems in which the radios can transmit or receive, but not both simultaneously. Another type of system is full duplex. In that case transceivers can receive and transmit messages simultaneously. Special provisions must be made to minimize the interference between the strong transmit signal and the weak receive signal. In an air/ground communications system this could conceivably be done by using separate frequencies for up and down links. If the frequencies are well-separated it may be possible to construct filters capable of reducing interference to an acceptable level. However, the amount of filtering required is considerable.

As an example, consider a scenario where a receiver is receiving a signal from a 10 watt transmitter at a distance of 160 nmi. At a frequency of 125 MHz the received power is nominally -84 dBm. Now suppose the receive platform is simultaneously transmitting a 10 watt signal at another frequency through a separate antenna. On a single platform it is seldom possible to achieve more than 40 dB isolation between antennas so that the unwanted interfering signal will be at a level of about 0 dBm at the receive antenna. If a signal-to-interference ratio of at least 10 dB is required then at least 94 dB of isolation is necessary. It is likely that expensive transmit and receive filters and careful attention to transmit waveform shaping would be required to implement a pure FDM full duplex system; i.e., pure FDM and full duplex may not be a good match. (As discussed further below, the transmit/receive isolation problem can be mitigated by separating transmit and receive frequencies into two well-separated bands).

Code Division Multiplexing

In a pure FDM system the approach is to make the individual voice channels as narrow (in frequency) as possible so as to pack as many as possible into a given 25 kHz channel. For code division multiplexing (CDM) the opposite is true. In this case each channel is spread over as much of the frequency band as is practical. This spreading can be accomplished by increasing the instantaneous bandwidth or by changing the carrier frequency of a narrowband signal either periodically or aperiodically. The first technique is often called direct sequence pseudonoise (DSPN), or simply spread spectrum. The second technique is usually referred to as frequency hopping (FH).

Consider the DSPN approach. In this case a digital signal is spread in bandwidth by overlaying the relatively low rate information bit sequence with a "chip" sequence whose chip rate is much higher. The instantaneous bandwidth is roughly equivalent to the number of chips per second. In pure CDM all users share a common frequency band, and different signals are separable because they have orthogonal (or nearly orthogonal) chip sequences. The chip sequence is the "code" referred to in CDM. In general, if a radio is receiving one CDM signal, all the others that occupy the same band act as noise; however, the noise contribution is reduced by the spread spectrum processing gain which is approximately the ratio of chips per bit.

There are a number of good reasons to consider DSPN. One primary benefit of such a scheme is that it is potentially more efficient than FDM in some circumstances. In the DSPN case interference is generated only by the transmitters which are "on" at any given time. Thus, capacity estimates can be made based on the expected peak activity (in a geographical area) for the system as a whole. In the FDM approach each set of users (e.g., a controller and aircraft under his or her control) is given exclusive rights to a frequency channel. This channel is removed from the pool of frequencies for other users whether or not it is actually in use at a given time. Thus, to the extent that the channels are not in constant use, there is a good deal of unused capacity in the FDM system. Everything else being equal, the CDM approach allows for more system participants, provided that the individual users' duty factors are low.

Another advantage of DSPN which is sometimes useful is its resistance to frequency selective multipath fading. Because of the wideband nature of the signal, the effect of narrowband disturbances is reduced. This feature may be very useful in land-mobile communications where over-the-horizon effects make such fading prevalent. However, in the ATC application the desired links are well-above the horizon or just above the horizon (nap-of-the-earth). In the former case, fades are generally not too deep; and in the latter case, fading can be deep but is wideband in nature. Thus, the antifading properties of DSPN, while not irrelevant, are not an overriding factor in the decision process.

Another advantage of DSPN is its potential to coexist in a relatively benign way with the current system of air/ground communications for ATC. Because the power of an individual transmission can be spread out over a relatively wide band, the power interfering with an individual narrowband AM signal can be quite low. To quantify this notion, consider the following example. Suppose that the information to be transferred is a 6.4 kbps voice signal. (The rate 6.4 kbps is chosen as an example only for the sake of consistency with the previous example. This rate is <u>not</u> being advocated here as the preferred solution). In principle this could be spread over the entire band from 118 MHz to 137 MHz. However, the part of this band actually allocated to the ATC voice communication function is broken into three discrete bands. The widths of these bands are approximately 4 MHz, 5 MHz, and 5 MHz as shown in figure 10-1. In order to fit within the smallest band the chipping rate could be 2.56 MHz. The processing gain of such a system would be 2.56 MHz/6.4 kHz = 400 which is equivalent to 26 dB. (Perhaps 20-24 dB is a more realistic range of numbers if practical implementation losses are taken into account.) Now the fraction of power transmitted by a DSPN transmitter which finds its way into the audio pass band of an AM radio is just 7 kHz/2.56 MHz = 0.003 or, equivalently, -25 dB. The rejection of unwanted signals for AM and DSPN receivers is similar because the information bandwidths are similar. The processing gain of a DSPN system seems quite high; whether it is high enough will be discussed below.

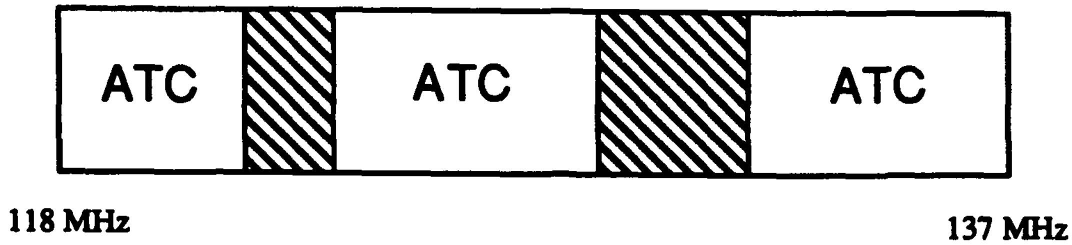


Figure 10-1. Air Traffic Control Radio Spectrum

Before continuing to discuss in more depth the advantages and disadvantages of DSPN some important points should be emphasized. The first point is that DSPN is an intrinsically digital process and so it really makes sense to use it only if the underlying information is digital, i.e., digitized voice is implied. Second, because of the way the frequency band is divided, a certain amount of FDM is unavoidable. If the chipping rate is 2.56 MHz then the band could be divided into four or five separate segments. This is not necessarily a deficiency; it just means that "pure" DSPN is not a good solution.

One disadvantage of DSPN is the "near/far" problem. One manifestation of this problem occurs at ground receivers. For example, suppose that two airborne transmitters are transmitting to the same ground site, and suppose the range of one is 10 nmi and the second is 160 nmi. The ratio of the powers of the two received signals is $(160/10)^2 = 256$ (i.e., 24 dB). The receiver for the weaker signal has a processing gain of 24 dB, say, so that the received signal-to-noise ratio is 24 dB-24 dB = 0 dB. This is typically <u>not</u> enough to provide adequate performance. There are, however, some ways to alleviate this problem. One possibility is adaptive power control. This method provides for the reduction in power of short-distance links and is accomplished by providing an uplink beacon so that the airborne radio can estimate received power. If the links are reciprocal then the airborne radio can transmit just enough power to provide good performance. Ideally, all signals arrive at the ground site with equal signal strength. A second technique to ease the near/far problem is the use of adaptive antenna arrays. These systems provide adaptive antenna patterns which reduce gain in the direction of strong signals. They tend to equalize received signal power except in cases where the two interfering signals are in the same direction.

A similar near/far problem can occur in airborne receivers. An aircraft flying in the vicinity of one ground transmitter while trying to receive from a more distant ground transmitter can experience high levels of interference which may stress the limits of the system's processing gain. In this case adaptive power control is not a viable solution since the ground transmitter may be transmitting to numerous receivers (so there is not a unique optimal power level). Adaptive arrays might work but they would be prohibitively expensive to place on aircraft. A third solution would be to limit the maximum range of the links. This solution would require that aircraft change their ground sites often as they fly across a geographic region. System overhead would have to be devoted to maintaining connectivity to the correct sites.

A third near/far problem can occur if an airborne transmitter is operating in the vicinity of an airborne receiver tuned to a different channel. This situation is similar to the one described in the previous paragraph. In this case, an additional means of addressing the problem is to use separate frequency bands for up and down links. If this is done then the airborne receiver can use frequency selectivity in addition to processing gain to reject unwanted signals.

Note that the use of separate up and down frequency bands can be used to address full duplex operation if that is a requirement. The interference rejection requirements for full duplex operation were discussed in the FDM section. While it may be difficult to provide the filtering necessary for full duplex in the case of separate up/down-link frequencies, it is clear that it would be impossible using CDM alone.

The reason why the near/far problem is so difficult for the pure CDM approach is that the amount of rejection of unwanted signals is set by the processing gain which is limited by available bandwidth. For the case of FDM the unwanted signal rejection is more dependent on frequency separation and by the complexity of the radios' filters. The first area can be addressed by frequency management and the second area just depends on details of the waveform and on radio design.

One final trouble area associated with DSPN exists in cases where the spread spectrum system coexists with a narrowband system (e.g., the current AM voice communications). An additional near/far problem arises if a nearby AM transmission interferes with the reception

of a distant DSPN signal. As discussed previously the processing gain of the DSPN system may not be enough to provide an adequate signal-to-noise ratio. Adaptive power control is not a viable solution since the current AM radios have no provision for such control. Adaptive arrays might help, but could be prohibitively expensive (especially in airborne installations). A technique which can help is known as "excision". In this method the incoming signal is digitized and Fourier transformed. Narrowband interferers appear as peaks in the resulting waveform; these peaks are then "excised". The resulting waveform is inverse transformed and processed as an ordinary DSPN input. This technique can be very effective, but makes intensive use of digital signal processing resources and will add to the cost of the radios. (Note, however, that the cost of high-speed processing chips continues to decrease.)

An entirely different sort of CDM technique which spreads an individual communications link over a wide frequency band is frequency hopping (FH). In this method, the instantaneous bandwidth is small, but the carrier frequency is changed from time to time (either periodically or aperiodically) so that an individual link can eventually make use of the entire available spectrum. There are two important choices which determine the performance of a frequency hopping systems. The first is the frequency hop pattern, i.e., the set of frequencies that will be hopped over and the sequence in which the frequencies will be used by a particular communications channel. The second parameter is the frequency hopping rate.

The choice of the frequency hopset is determined by whether or not the FH system will share frequencies with the current AM system. Whether or not it is acceptable to share frequencies depends on the mutual interference between the two systems. The effect on the current system will be minimized if the hopping rate is fast enough so that an individual hopping pulse does not break squelch; however, an individual hopping pulse may be "heard" by a receiver whose squelch has already been broken. This interference may be acceptable if the number of interferers is small enough so that an active link is not "hit" too often. Conversely, the effect on the hopping system will be minimal if the number of active AM radios is not too high at any given time. The questions of the correct hop rate and what constitutes an acceptable interference level will be discussed below.

Whatever hopset is chosen, a choice has to be made as to the type of hopping pattern to be used. One choice is to ensure that each channel (or circuit) chooses a different frequency at any given time. This is known as an orthogonal hopping pattern. The second possibility is to allow different nets to hop independently so that occasional collisions are possible. The self-interference generated will be tolerable if, once again, the number of interferers is not "too high".

An appropriate hop rate can be chosen by considering certain performance and interference related constraints. For instance, if the hopping frequency set overlaps the set of frequencies used by the current system then the dwell time should be short enough so that squelch is not broken on the existing radios. The value therefore depends on the attack time of the squelch circuitry. As an example assume that a dwell time of 10 ms is sufficiently short. For a high duty factor waveform this corresponds to a hop rate of 100 hops per second. Note that if the FH technique is pursued further, the issue of the hop rate needs to be investigated in more depth.

Another contributor to hop rate selection is interleaver size. For digital approaches the effects of occasional interference bursts may be reduced by spreading the error bursts out in time. This is particularly true in data systems and certain voice coding techniques that employ error correction coding. For efficient error spreading it may be necessary to interleave the bits over 5 to 10 hops. However, the interleaving/deinterleaving process causes a throughput delay equal to twice the interleaver size. If we want the delay to be less than (say) 100 ms than the hop rate should be no less than 100 hops per second. Note also that a ten-hop interleaver will reduce an occasional burst of errors at 50% bit error rate (BER) to a more benign level (5% = 50%/10).

Note that analog signals could also be frequency hopped. In this case, the hops (i.e., frequency switching times) will cause noise bursts in the reception. For a 100 hps system these bursts will occur at a 100 Hz rate. Such an interruption rate may allow for an intelligible signal; however, message quality would probably be reduced significantly. Lower hop rates may result in higher quality voice transmission if the hop rate components in the audio signal are subaudible. On the other hand the associated longer dwell times would probably be unacceptable from a mutual interference standpoint.

From the above discussion, it may be apparent that only certain combinations of hopset, hop rate, etc. make sense. For instance, analog voice transmission may require a slow hop rate. This in turn may require that the hop set be disjoint from the AM frequency set and orthogonal. It can then be shown that the resulting capacity is no greater than that of a non-hopping FDM system. The added expense of FH would be wasted. The real gain for FH lies in the possibility to coexist with the existing AM system and to use non-orthogonal nets which will experience tolerable levels of interference.

We can estimate the level of interference by assuming (once again) a 6.4 kbps voice system. If we assume that the FH system hops on 5 kHz centers then there are about 2800 hopping frequencies in the allocated band. An individual interferer may occasionally cause a burst of errors for one hop consisting of 64 bits (at 100 hps). A strong interferer will create (on the average) 32 errors; however, a ten-hop interleaver will spread these over 640 bits so that the "peak" error rate is only 5%. The "average" error rate is just (1/2800) (1/2) = 0.02%. If a 5% BER were tolerable then (roughly) 250 radios could be transmitting in close proximity before unacceptable levels of interference arose.

The performance of FH can be compared in a qualitative way with that of DSPN. Since both techniques involve the sharing of spectrum resources they both must tolerate certain types of self-interference. In DSPN this interference manifests itself as a rise in the noise level whose magnitude depends on the number and location of the interferers. In FH, on the other hand, the interference consists of bursts of noise. The effect of this noise depends more on the number of interferers than on their geographic locations. In this respect, the FH approach has better resistance to the near/far problem than DSPN because the maximum effect of an individual interferer is limited by the probability of frequency overlap and not by signal strength. Another (perhaps small) advantage of FH over DSPN is the fact that the FH approach can allow a single channel to hop over all available frequencies, whereas an individual DSPN must confine itself to one of the three disjoint frequency band available to ATC.

The ability of a pure FH system to support full duplex operation is limited by stringent filtering requirements as outlined previously in the section on FDM. The problem is particularly difficult since the filters would have to "hop" along with the frequency hopping pattern. An alternative is to divide the available spectrum into two disjoint bands and use one

band for uplinks and one for downlinks. This approach may render the filtering problem tractable, although not easy. Note that for all the approaches mentioned so far (FDM, DSPN, FH) full duplex would require the radios to be tuned to two frequencies simultaneously and would require two frequency synthesizers and other duplicate parts in addition to high performance filters.

A third general approach to CDM (besides DSPN and FH) that has been proposed is a "chirp" waveform. Chirp techniques are most common in radar applications, but they can be used for communications. In this case the signal is spread across a wide bandwidth by linearly varying the carrier frequency across the band during each bit interval. A digital "1" may be designated by an increasing frequency, and a "0" by a decreasing frequency. The properties of the chirp technique are similar to those of DSPN. A potential drawback of this technique is the difficulty in implementing it. One suggested implementation uses surface acoustic wave (SAW) devices to match the chirped signal. However, if the bit rate of the signal is less than about 10 kbps then the length of a filter matched to a bit duration (100 μs) would be over a foot long. Such a physically large device might be difficult to fabricate cheaply. Other technologies, such as analog charge coupled devices (CCDs), might be applicable to a chirped waveform; and a more detailed tradeoff analysis would have to be done to determine if "chirp" is a viable alternative. A preliminary estimate indicates that this approach is not very promising.

A final approach to CDM is known as time hopping (TH). In this method, which is intrinsically digital in nature, the instantaneous bandwidth of a signal is increased so that the waveform duty factor can be reduced substantially below unity. In this case the transmitter must transmit only a (small) fraction of the time. If the actual transmission time is pseudorandomized according to which TH channel a radio is on, then this time pattern is the "code" referred to in CDM. The properties of TH are very similar to those of FH. Indeed, TH and FH are very often used in conjunction with one another in hybrid systems. However, TH requires high peak power and may not be compatible with low-cost solid-state transmitters. It does not seem that TH offers any advantages over other CDM techniques.

Time Division Multiplexing

The last major approach to multiplexing which will be covered in this memo is time division (TDM). TDM is similar to TH except that in TDM the transmission times associated with a particular channel (or circuit) are usually assumed to be regular (or periodic). The number of circuits which can share a particular 25 kHz channel depends on the information rate and the bandwidth efficiency of the chosen waveform. To continue with the hypothetical example used previously, assume that the voice bit rate is 6.4 kbps. If the waveform has an efficiency of 1.6 bits per second per Hz then 4 kHz of bandwidth is needed for a 100% duty factor waveform. If the duty factor were reduced to 1/6 then the instantaneous bandwidth required would be 24 kHz. While the above calculation indicates that up to 6 voice channels can be multiplexed onto a single 25 kHz channel, it does not include any provision for synchronization, guard times, and overhead for channel maintenance. It is likely that less than 6 channels (possibly 4 or 5) could fit into one 25 kHz allocation. Note that this calculation is very similar to the capacity calculation for FDM. This is because the two approaches are nearly identical if one exchanges the frequency domain for the time domain.

In spite of the close relationship between FDM and TDM, there are some very definite advantages to TDM under certain conditions. If the system concept includes full duplex operation then transmit and receive times can be separate so that a single transceiver can service both up and down links. This removes the requirement for duplicate frequency synthesizers, etc., making for a simpler radio design. Note, however, that this type of radio must rapidly switch from transmit to receive (and vice versa). Also, if the up and down link frequencies are different then the synthesizer must be able to change frequencies relatively quickly. The required cycle time of the TDM network depends primarily on requirements for a minimal voice latency period (or access time). That is, if the maximum allowable access time is 100 ms, then the time between transmit (or receive) time slots should be no greater than 100 ms. Note that if a TDM approach is used then additional features can be added to the system relatively easily. For instance, in a system with a separate service channel, time slots can be set aside for service channel use on a noninterfering basis.

In this section the major multiplexing techniques have been discussed in the context of their application to air/ground ATC communication. To a large extent, the techniques (FDM,

CDM, and TDM) have been discussed separately. However in many cases the best solutions are hybrids of two or more techniques.

10.4 TRADEOFFS

Some of the criteria which would allow one to choose between the various types of multiplexing were described in the previous section. In this section the criteria are discussed systematically along with the tradeoffs they engender.

A. Voice modulation: An important factor in the choice of multiplexing used is the type of voice modulation. The main decision is between analog and digital techniques. If any type of analog voice is chosen then most types of CDM and TDM are ruled out. DSPN and chirp are intrinsically digital techniques; and time hopping and TDM manipulate time in a way that is unnatural for a continuous time analog signal.

There is some possibility that frequency hopping could be used with analog modulation; however, as discussed in the section on FH, there may not be an appropriate hop rate for this application. If the hopping pattern is not orthogonal and/or overlaps with the allocation for the current fixed frequency system then the hop rate would need to be relatively high. A hop rate of at least 100 hps would be required for interference reasons, but would not provide good quality voice. If the hopping pattern were orthogonal and non-overlapping then the hop rate could be slowed down; but in this case there would be no real advantage over a nonhopped FDM system. Added complexity would provide no commensurate benefit.

In summary, FDM is the only appropriate multiplexing technique for analog voice. For digital voice, any of the multiplexing techniques is possible. The choice would be based on other criteria.

B. Support of special features: Some special features that a system may or may not require include full duplex operation and/or the simultaneous monitoring of a voice channel and a service channel. (Service channels are an important feature of systems with automatic frequency handoff, etc.) In a pure FDM system each simultaneous function provided needs separate hardware to some extent. For example, for full duplex operation there need to be

separate frequency synthesizers for the transmit and receive frequencies. Also, there needs to be a great deal of filtering to ensure that the received signal is not overwhelmed by the transmitted signal. This filtering requirement can add to the complexity of the system, although it may be simplified somewhat if up and down links are confined to separate bands with guard bands in between.

In a pure DSPN system the problem is similar. In this case the unwanted signal rejection provided is limited by the number of chips per bit. Again, the problem can be alleviated to some degree by separating up and down links into separated frequency bands. If a more complex system is desired (with multiple up and/or down links), the various up links (and down links) can be distinguished by using different PN codes so that no more than two synthesizers would be necessary.

Supporting full duplex operation is simpler in the case of TDM. Only one transceiver is necessary. The radio can support uplinks and downlinks on a time shared basis. Since a radio is never trying to transmit and receive simultaneously, self-interference is essentially eliminated. However, because each function has a duty factor less than one, latency periods depend on the time it takes to cycle through these functions. In other words, the time between push-to-talk and transmission depends on the time slot cycle time. This time needs to be kept to a minimum (say, less than 100 ms) so that the beginning of a voice message is not "clipped" and/or the throughput delay is not perceptible.

Note that the discussion of the previous paragraph refers to a platform with a single ATC radio. On the ground, multiple channels may be supported by a single radio site. In such a case it may not be possible to separate transmit and receive times. On the ground, however, transmit and receive antennas can be separated to achieve greater isolation than on an aircraft. Nevertheless, it may once again be necessary to use separate up and down link frequencies to minimize interference. Separating the up and down link frequencies into separate bands might also be helpful.

C. Communications capacity: The primary goal of a system upgrade is to increase system capacity, i.e., to maximize the effective number of voice circuits, or channels, within a given geographical area. The capacity depends in a complicated way on a number of variables. Important factors include the size of the coverage volume (particularly the extent of altitude

coverage), channel usage statistics, multiplexing scheme, tolerable interference level (or BER in the case of digital modulation), spectrum allocation, and information bandwidth.

It is useful to have a specific scenario in mind when comparing the various multiplexing methods. It will be assumed that all users are, at any given time, assigned to specific user groups which consist of a controller and the airborne users under his or her control. Each controller has the exclusive use of one of more communication circuits which are connected with specific radio sites. Airborne users periodically change sites, circuits, and/or controllers as they fly from place to place. These changes may be manual or automatic depending on the details of the system operation.

The radio sites are assumed to be arranged in a somewhat regular pattern throughout the relevant geographic region. In previous work [4, 5] it was shown that any volume of space ranging from altitude 1125 ft. up to 72000 ft. could be covered by a set of radio sites each of which was separated from its three nearest neighbors by 40 nmi. Within the context of TDM or FDM this configuration would allow for double coverage of any aircraft in the volume by populating the sites with 20 sets of uplink frequencies and 20 sets of downlink frequencies.

Note that certain assumptions were made in setting up this example. For instance, the altitude range could be changed [6], double coverage (for emergency backup) can be provided in number of ways, and separate up and down link frequencies need not be used. However, the given case serves as a good basis for comparison. It is further assumed that the system uses 6.4 kbps digitized voice which can tolerate a 5% bit error rate.

Approximately 14 MHz of spectrum is allocated for the function under consideration (see figure 10-1), which equates to about 560 25 kHz channels. In a previous section it was estimated that each 25 kHz channel could support approximately five 6.4 kbps voice channels if the system used TDM or FDM. In either case the number of channels is about 2800, of which 1400 can be used for uplinks and 1400 for downlinks. These frequencies must be divided into 20 sets each to allow for efficient frequency reuse. Each set would have approximately 70 channels pairs or circuits. This means that each radio site could support the activities of up to 70 user groups (i.e., controllers).

In the case of a DSPN system all the assumptions remain the same except for the frequency reuse pattern. It will be assumed that the spectrum is divided into four bands, each approximately 2.5 MHz wide, with two supporting uplinks and two supporting downlinks. As described previously each link could theoretically enjoy a processing gain of 26 dB. To allow for system losses and overheads a more realistic estimate of processing gain is taken to be 24 dB. Since there are two uplink frequency bands, it is assumed that alternate radio sites use alternate bands in order to minimize interference.

In the specified scenario, any aircraft at elevation 1125 ft. has two radio sites within line of sight. As the aircraft increases in altitude the number of radio sites within view increases until about 150 can be "seen" at 72000 ft. The accumulated power transmitted at these numerous sites can add up to a very noisy environment. It is possible to analyze this situation fairly precisely to show that (in the worst case) at 72000 ft. the signal-to-noise ratio would be -8 dB if there were one transmitter at each site and the aircraft were "listening" to the nearest available transmitter site. Assuming that a signal-to-noise ratio of 4 dB is required for an acceptable BER (using differential phase shift keying [7]) the system margin would be about 24 dB - 8 dB - 4 dB = 12 dB. This means that about 16 transmitters could be in operation at each site before unacceptable performance degradation began.

Note that the 16 circuits per site for the DSPN case is significantly lower than the 70 circuits per site for the TDM and FDM cases. The difference in performance is ultimately attributable to the limited processing gain of the DSPN system. These capacity estimates do not tell the whole story, however. In the case of DSPN the number represents the number of radios actually transmitting at a site at any given time. Thus, if duty factors are low, more circuits can be assigned to a site. Analysis of the example presented above shows that if the average duty factor is less than about 20%, then more than 70 circuits can be assigned to a site using DSPN.

Another advantage of DSPN is that there appears to be more flexibility in distributing capacity on a geographic basis. In the TDM and FDM cases it is assumed that frequencies are assigned to sites in a coordinated, quasiperiodic way throughout the region of coverage. In the case of DSPN the assignment of circuits to sites is simpler because of the large number of pseudorandom codes available (the number of possible 400 chip codes is astronomical,

i.e., $2^{400} \approx 10^{120}$; even a subset of these with good orthogonality properties would be very large.). Thus, regions of high traffic density can be accommodated more simply.

One of the major advantages of DSPN lies in its potential to coexist with current air/ground communications system. In the calculations done above it was assumed that the same spectrum was available to all the systems. However, it may be that in the initial stages of deployment of a new TDM or FDM system many frequencies occupied by the current system would not be available for the new one. This could lower estimates of system capacity considerably. It should be borne in mind, however, that in the initial stages of deployment capacity requirements might be reduced.

A drawback of the DSPN approach presented here is that the airborne radio must maintain contact with the closest radio site in order to maximize the received signal-to-interference level. An aircraft flying at a speed of 400 nmi/hr might need to change ground sites approximately once every 12 minutes. These changes (whether they are done automatically or manually) would add to the communications traffic. In the TDM or FDM cases the frequency usage could be arranged in a layered, or tiered, way so that for higher altitudes, radio site changes would be needed much less often.

The previous discussion focussed on a comparison of DSPN versus TDM or FDM. Other CDM techniques such as FH or chirp could have been substituted for DSPN. The analysis would have changed, but the conclusions would have remained more or less the same. All the CDM techniques share the same advantages and disadvantages vis-a-vis TDM and FDM to a large extent. It should also be noted that the estimates made in this section are rough ones that involve various approximations and simplifications which, while good enough for comparative purposes, should be refined for any system given further consideration.

2. Self-interference: Self-interference is intrasystem interference experienced by one user of the system caused by all the other users of the system. In the last section on system capacity the discussion centered on interference to an airborne receiver due to a multitude of ground transmitters (uplink interference). A slightly different form of self-interference is caused by airborne transmitters (downlinks). Much of the problem for downlinks is similar to the uplink situation. An additional feature of the downlink problem was addressed in previous

sections as the near/far problem. Unlike the ground transmitters the airborne transmitters are free to move around. Thus, it is possible that a receiver could be attempting to demodulate a message from a distant transmitter while experiencing interference from a nearby one.

Since this phenomenon was discussed at length previously, only the conclusions will be repeated here. For TDM and FDM the different circuits are separated in frequency and/or time so that the problem is not a serious one. For DSPN and chirp systems the unwanted signal rejection is limited by the processing gain so that the problem can be (partially) solved only by ensuring that the situation does not arise very often. This can in principle be done by requiring that airborne radios always be "connected" to the nearest radio site. This might require frequent site changes. For FH (and TH) systems, the downlink interference is statistical in nature. The near/far interference can be held to tolerable levels if the number of hopping frequencies is high enough and if the voice (or data) coding technique can tolerate some specified error rate.

E. Intersystem interference: It is assumed that any new system will have to coexist with the current air/ground communications system for an indefinite period of time. The new system should not cause levels of interference which would significantly degrade the current one.

For TDM and FDM interference can be avoided only by assigning the new system frequencies which are currently not used. This might limit the capacity of the new system (initially, at least). For the CDM systems it is possible that the old and the new systems could share spectrum resources. For DSPN systems spectrum sharing may cause unacceptable interference if traffic levels are high and/or if geometrical considerations cause near/far problems between the two systems to arise frequently. Note that narrowband interference to a DSPN radio can be mitigated by excision. Such an option does not exist for interference to a narrowband radio. The other types of CDM systems perform better in this respect because they will occupy an individual common frequency for only short periods of time. Thus, a single transmitter cannot "wipe out" a received signal under any circumstances.

F. Fading resistance: While the need to take fading into account is important in the design of any communication system, it is not a major factor in this decision process. This is so because the links involved are assumed to be (almost always) line-of-sight links. Analyses have shown that for VHF air/ground scenarios most fades are significantly less than 10 dB

and that this amount of fading can be built into the system margin. Fading due to specular reflection near the radio horizon can be much deeper; however, it is very broadband in nature so that spectrum spreading techniques do not offer much relief. Thus, to a large extent fading considerations are not a major contributor to the multiplexing decision. Note that this does not mean that fading cannot be a major consideration in other decisions, such as the one between amplitude and frequency (or phase) modulation.

G. Complexity of design: In the previous sections the emphasis was on the capabilities of the various techniques to meet system requirements. It is also important to consider the design complexity implied by the different options since complexity can ultimately be translated into cost. Also, a complex design which could be implemented by only a small number of "high-tech" manufacturers might not be acceptable. The relative complexity of the equipment depends to a large extent on the features it will have to support. The major consideration is whether the system will support full duplex or half duplex operation.

If half duplex is desired then, by definition, the radios must be able to transmit and receive, but not at the same time. In this case complications due to stringent transmitter/receiver isolation requirements are not necessary. TDM and FDM implementations are roughly equal in complexity. For FDM selectivity requirements are greater than for TDM, while for TDM signal processing speeds and synchronization requirements are greater. If the radios must retain backward compatibility with the current AM system then there may be an additional advantage to a TDM solution because common IF filtering could apply to both the old and the new modes.

The CDM solutions would certainly be more complex. For a DSPN system, signal processors would have to operate at the chip rate and not the bit rate. Also, it may be necessary to implement adaptive power control and excision in order to control self-interference and intersystem interference. For a chirp system complex analog (SAW) correlators would be necessary. FH and TH waveforms would require fast tuning synthesizers and perhaps complex synchronization circuitry.

If full duplex operation is desired, then the picture changes somewhat. TDM has a definite advantage over the other techniques because full duplex operation can be implemented by rapidly changing from transmit to receive. An individual radio never has to

do both simultaneously if the transmit/receive cycle time is short compared to an acceptable voice throughput delay. For all the other techniques transmit/receive isolation becomes a major consideration. The problem can be mitigated to some extent by transmitting up and down link messages in well-separated frequency bands. Nevertheless, this amounts to an added complexity for all but TDM solutions. Besides the filter problem, non-TDM solutions also suffer from a requirement to replicate hardware that does not exist for TDM. For instance, separate synthesizers will have to be implemented for the up and down links. Note that if there is a requirement to communicate on a service channel in addition to a voice channel, the difference between TDM and the rest becomes even more pronounced.

The information in this section is summarized in table 10-1. The reader should be aware that this table, like most tables, tends to oversimplify the issues. Nevertheless, the table may be useful in putting the issues into a manageable perspective.

10.5 IMPACT/IMPORTANCE

While the ultimate goal of an upgrade of the VHF air/ground communications system is to increase the system capacity, two other factors play pivotal roles in deciding which multiplexing options are viable ones. These factors are the choice ot voice modulation and the choice between half and full duplex.

If analog voice modulation is chosen then the preferred type of multiplexing is FDM. All other techniques are either intrinsically digital or require parameters which are not convenient for voice communications. If the voice coding is digital then all of the multiplexing options are available.

If half duplex operation is the standard operating mode of the system then any of the multiplexing ideas is viable. On the other hand, if full duplex is required then TDM has obvious advantages over the other techniques. If each of the required up and down link functions can be given its own time slot then problems of transmit/receive isolation and equipment complexity are mitigated. This does not necessarily mean that the other techniques do not have a role to play in this case. Within a time slot any of the other techniques may or may not be applicable; e.g., a hybrid CDM/TDM system may be a viable

option which combines the good features of both techniques. A hybrid FDM/TDM system might also be useful.

Table 10-1. Comparison of Multiplexing Techniques

	FDM 1)	TDM 1)	CDM 1)
Voice Modulation	Analog and Digital	Digital	Digital
Full Duplex Support	Moderately Difficult	Simple	Moderately Difficult
Capacity 2)			
High Duty Factor	High	High	Moderate
Low Duty Factor	High	High	Highest
Self-Interference	Low	Low	Moderate
Intersystem Interference			
Separate Frequencies 3)	Low	Low	N/A
Overlapping Frequencies 2)	High	High	Moderate
Fading Resistance	N/A	N/A	N/A
Complexity	Low/Moderate	Low/Moderate	High

1) All systems assume separate uplink and downlink frequencies

2) Assume systems use full 14 MHz allocation

3) Assume systems use frequencies not used by current ATC system, but within 118-137 MHz band.

A system which meets all other requirements should then be judged according to how much it increases the capacity of the system. The capacity is influenced by uplink interference and downlink interference considerations. These are treated separately due to different geometric considerations. Whereas the up link transmitters are fixed (ground radio sites), the down link transmitters are mobile (aircraft). Because of their mobility these transmitters can cause a special problem —the near/far problem — for CDM techniques. For FH and TH systems the problem is not too severe since it depends on transmit duty factor

more than on relative geometry. For true spread spectrum techniques (DSPN or chirp) special measures may be required. These include adaptive power control and/or adaptive antenna arrays.

The ability of any new system to coexist with the current ATC voice system is also very important. For TDM and FDM systems coexistence can be accomplished by arranging to have the new system utilize frequency resources unused by the current system. Initially, at least, this will limit the capacity of the new system because most of the spectrum is already in use. However, as more and more operators use the new system, more frequencies can be allocated to it. Note that to implement this type of evolution it may be necessary to redistribute frequency usage periodically.

The CDM approaches are simpler from an evolutionary point of view. They are designed to share spectrum resources by allowing a certain amount of intrasystem and intersystem interference. This is possible if the resulting interference is kept below acceptable levels. In previous sections it was suggested that it is possible that different CDM systems might be able to coexist with the current system. <u>This conjecture would need to be verified in much more detail if a CDM approach is considered likely</u>. A strawman system should be designed in order to allow a more "in-depth" analysis of such issues as the actual processing gain available, quantitative estimates of acceptable interference levels, and costs associated with adaptive power control, excision, and adaptive antenna arrays (if any).

Resistance to fading is another phenomenon which is sometimes used to judge system alternatives. However, in this case fading resistance might not provide much help in deciding between FDM, TDM, and CDM since there are not large differences in fading performance among them. CDM has some advantages against frequency selective fading. However, it appears that the fading experienced at VHF is either wideband (at the horizon) or not exceedingly deep. In addition, the bandwidth provided (particularly for DSPN) is just not wide enough to have a major effect. Thus, while the fade resistance provided by CDM would be helpful, it is probably not large enough to weigh heavily in a system decision.

A final factor affecting a decision between system alternatives is equipment complexity. Complexity is important since it is ultimately reflected in system cost and

system reliability. Although actual cost and reliability estimates will not be made here, certain assertions pertaining to complexity can be made.

If the system design requires only half duplex operation then FDM is probably the simplest approach. Complexity issues include the need for increased selectivity (i.e., increased filter requirements) and enhanced frequency stability. An FDM approach also is compatible with analog as well as digital voice transmission. TDM and CDM are more restrictive in that they apply primarily to digital voice systems. TDM may be slightly more complex than FDM due to increases in processing speed requirements. The high speed is necessary to address higher baud rates and synchronization requirements. On the other hand, a TDM system may be able to take advantage of some common hardware simplification if the new radio retains a mode which is backward compatible with the current AM voice system. In particular, IF filtering may be simpler if both modes require the same IF bandwidth (25 kHz).

The CDM systems would be the most complex. If DSPN is employed then the processing would have to take place at the chip rate rather than the bit (or symbol) rate. For FH, hopping frequency synthesizers would increase complexity. For a chirp system SAW filters may prove to be prohibitively expensive. TH would incorporate the complexities of both TDM and FH.

If full duplex operation is required then TDM has certain obvious advantages. With TDM a single radio never has to transmit and receive simultaneously. Thus, a single radio needs only one frequency synthesizer and does not need to duplicate a variety of other parts. Also, transmit/receive isolation requirements are eliminated. Thus, the most economical type of full duplex system is probably one that includes TDM, either alone or in combination with one or more of the other techniques.

If, for some reason, TDM is not applied to a full duplex system then the hardware duplication and isolation issues will have to be addressed. The isolation issue can be alleviated to some extent by separating the transmit and receive functions into two well-separated bands. This may be convenient since the frequency allocation for ATC is already segmented; but putting up links and down links into different bands will require extensive changes to current frequency assignments.

Note that the use of separate bands for transmit and receive (or, equivalently, for up and down links) may be a good idea even in a system employing TDM. Because ground radio sites will undoubtedly have to support multiple, independent links, it may not be possible to separate transmit and receive functions in time. Therefore, a certain amount of isolation will be required. On the ground, antennas can be spatially separated to provide some additional isolation; but separate transmit and receive bands will simplify the antenna placement problem. Also, in the case of multiple uplinks on multiple frequencies intermodulation products (IMPs) can be generated by hardware nonlinearities. If the transmit frequencies are grouped together then the IMPs will tend to be located in the vicinity of this group. Thus, they will not interfere with reception. Also, cross modulation and receiver desensitization problems will be alleviated by separating transmit and receive frequencies.

10.6 CRITERIA FOR DECISION

There are four major criteria for judging the advantages and disadvantages of a new system concept. These are listed briefly.

1. Capacity: The system should maximize the system capacity. For a given distribution of radio sites this equates to providing the maximum number of effective voice circuits per radio site.

2. Support of special features: The overall system may or may not incorporate special features such as full duplex operation and/or the monitoring of service channels. The multiplexing system should be able to support all the required features.

3. Evolutionary implementation: The system should provide for coexistence with the current system and a gradual phasing in of the new system. It may be that the old (current) system is never removed entirely. The new system should also incorporate features which will allow it to coexist with any new aspect of the ATC system, including increased use of data communications.

4. Cost: The new system should satisfy the above criteria in a cost effective way which minimizes one-time cost and life-cycle cost for the airborne radios and the ground

infrastructure. The equipment suite should include an airborne radio which is affordable by the general aviation community.

In light of these criteria and the issues and tradeoffs discussed previously, certain general statements can be made.

1. If analog voice is used then FDM is the preferred multiplexing technique.

2. Using separate up and down link frequencies in separate frequency bands provides at least three benefits. These include enhanced transmit/receive isolation, improved IMP (and cross-modulation and desensitization) performance for collocated transceivers (i.e., at ground sites), and more efficient frequency re-use.

3. If full duplex operation is required, then some form of TDM should be included in the system design. Other techniques such as FDM and/or CDM may or may not be used in conjunction with TDM. The optimal mixture may depend on the features desired, e.g., full duplex for both voice and a service channel will put some restrictions on the possible options.

10.7 CONNECTIVITY/RELATIONSHIP WITH OTHER DECISIONS

This paper is focussed primarily on the branch of the decision tree labelled "multiplexing", 4.3.2. This branch of the tree is related to branch 6.2 which concerns itself with efficient frequency reuse because the reuse pattern will reflect itself on the number of frequencies available at any given antenna site (see also branch 4.3.1).

This topic is also related in a general way to branch 4.2 which deals with channel access. In a sense, the multiplexing technique creates a number of communications channels and the access technique governs the way these channels are used.

It is clear that this topic is also related to branch 2.2.1.1 on voice quality. As stressed numerous times in the discussion the type of voice modulation has a profound effect on the types of multiplexing possible and on the number of circuits which can be packed into a 25 kHz channel.

Finally, this topic is related to decision tree branch 2.3 which deals with decreasing operator workload. The various means of reducing workload are connected to the "features" which the multiplexing scheme must support. These in turn generate the technical requirements which the system must meet.

SECTION 11

RANDOM ACCESS

11.1 CONTEXT

A systematic process for examining technical alternatives for improved air/ground communications in air traffic management has been established. A decision tree structure is shown in Appendix A that attempts to organize various alternatives in a top-down hierarchy. This provides a framework for evaluating potential solutions that can be represented by paths through the decision tree.

This paper addresses a particular set of technical tradeoffs in the decision tree, namely, the random access branchpoint 4.2.3 of the communications throughput subtree. The organization of this paper is as follows: Background, Issues, Tradeoffs, Impact/Importance, Transition, Criteria for Decision, and Connectivity/Relationship with Other Decisions.

11.2 BACKGROUND

This paper focuses on a generic class of communications channel access techniques called random access (RA). This generally means that many small duty factor users attempt to gain access to the channel whenever they have a need, without prior arrangements for dedicated channel access as with fixed assignments (node 4.2.1 of the decision tree) or without requesting shared access from a controlling authority as with demand assignments (node 4.2.2). Small duty factor means that each user generates new messages at a low enough rate that these messages would occupy the channel only a small fraction of the time available if this user was the only one transmitting. The suitability of a RA protocol for a given application is mainly governed by the evaluation of the protocol's ability to generate high throughput with acceptable delay under conditions of "heavy" channel loading, i.e., when many users attempt to send a message over the communications channel. Throughput and delay are defined following equation (0) below. The protocol's stability, i.e., capability to recover "gracefully" from heavy loading conditions, is also an important factor. A protocol is said to be stable if there is some mechanism provided in the protocol which prevents the throughput from going to zero asymptotically as the channel loading increases

without bound. However, delay can become unbounded as the throughput increases even for a stable protocol.

Although some of the same principles of RA can be applied to analog communications, the paper is devoted to RA for digital communications only. That realm is not only where most of the theory has been developed but also where the bulk of the applications lie. Furthermore, we will restrict ourselves to single-"packet" messages. Although multi-packet messages may be attractive for various system applications, the essence of RA can be covered by treating single packets. A higher-level networking protocol, as opposed to a single channel data link access protocol, would address multi-packet messages. The main thrust of media access control for air/ground (A/G) communications can be handled by the data link layer, Layer 2 of the Open Systems Interconnection (OSI) model of the International Standards Organization (ISO) [1].

The duration of a typical digital message packet, T, equals the ratio of the number of information bits in the packet, L, to the information rate, R, in bits per second (b/s), i.e., $T = L/R$. The expected channel loading, G, is the average fraction of the time the channel is occupied by either newly generated packets or (old) packets that are being retransmitted, both from all users. Packets that did not get through to the intended receivers with sufficient reliability are the ones that need to be retransmitted. It is noted that if the application is packet voice, unreliable packets may be discarded without greatly affecting speech intelligibility. However, we will mainly concentrate on the packet data application here.

It is assumed that packets successfully received are acknowledged by the receiver sending an acknowledgment packet to the transmitter over a separate return channel without error. Furthermore, it is also (usually) assumed that the only packet errors that can occur on the forward channel are due to packet collisions, not any other sort of channel induced error.

For mathematical tractability most RA queuing models assume a Poisson distribution for packets "generated" within the network or "arriving" at individual user transmission buffers. (Sometimes this is taken to mean both new packets and old packets, i.e., packets that need to be retransmitted.) The probability that a total of k packets is generated or arrive in a time t is

$$P_k(t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \; k = 0, 1, 2, \; \cdots \qquad \text{(0a)}$$

if the generation/arrival rate is λ packets per unit time. A well-known property of this Poisson distribution is that both the mean of k, E(k), and the variance of k, σ_k^2, equal λt, i.e.,

$$E(k) = \sigma_k^2 = \lambda t. \qquad \text{(0b)}$$

Throughput, S, is the average number of transmitted packets that get through satisfactorily, i.e., that are received correctly at the intended destination receivers. If two or more packets overlap in time on the channel they are said to have "collided"; in this case it is generally assumed that none of the packets are received correctly.

Channel efficiency or utilization, U, can be less than the throughput if there is some overhead used in accomplishing successful packet transmissions. The mini-slotted alternating priorities protocol to be discussed in the issues section is an example of this.

Delay, D, refers to the average time it takes for a newly generated message to arrive correctly at the destination receiver. This expected delay monotonically increases, with increased channel loading. However, as we will see later, in RA schemes, delay is not necessarily single-valued with respect to throughput. In other words, as the channel loading increases beyond a certain point, throughput can decrease because of greater self-interference on the channel due to more retransmissions and the increased likelihood of packet collisions.

The delay requirement for a system can serve as an important discriminator among channel access alternatives. For example, in real-time voice conversations, an overall delay of no more than about 200 ms may be acceptable because such a short delay is almost imperceptible to a human listener. However, longer delays would become more noticeable and necessitate effort by the talkers to avoid interrupting each other. As many people have experienced over long distance telephone calls via satellite, one-way delays of 1/4 s or even more can be tolerable. The practical limit, even among professional operators, would probably be about 1/2 s [2]. The Federal Aviation Administration (FAA) is in the process of defining voice and data delay requirements [3]. Their NAS-SR-1000 document includes an A/G response time of 250 ms, an A/G transmission time of 6 s one-way, and minimal

channel throughput delay [4]. We shall assume 200 ms for the nominal Air Traffic Services (ATS) access delay requirement. Channel access schemes with an expected access delay much longer than that would therefore be deemed unacceptable.

ATS requirements are driven more by real-time voice needs since analog voice constitutes a major portion of present operational communications in Air Traffic Control (ATC). There is growing interest in data links [5] to complement voice. In the ATS realm data communications has the potential for improving the reliability (message integrity) of some voice communications and for decreasing the workload of pilots and controllers in areas of critical but relatively routine communications. For example, the many frequency handoffs presently required in most controlled flights could be accomplished automatically over a data link. Similarly, controller handoffs could also be handled semi-automatically under the cognizance and control of the operators involved.

Another realm of requirements is Aeronautical Operational Control (AOC) [6]. Here the push for improved data links is even stronger. For example, the Aircraft Communications Addressing and Reporting System (ACARS) provided by Aeronautical Radio, Incorporated (ARINC), and first introduced by United Airlines, has been highly successful and is growing in usage among the airlines. ACARS operates at 2400 b/s using the output of a Minimum Shift Keyed (MSK) digital waveform to modulate a standard AM radio [7]. The airlines want to increase the data rate to significantly alleviate the growing congestion on the ACARS channels. The Airlines Electronic Engineering Committee (AEEC) is currently developing a VHF Data Radio (VDR) which will operate at either 21 kb/s or 42 kb/s depending on whether the radio is using 4-ary Offset Quadrature Amplitude Modulation (4-OQAM) or 16-ary OQAM (16-OQAM).

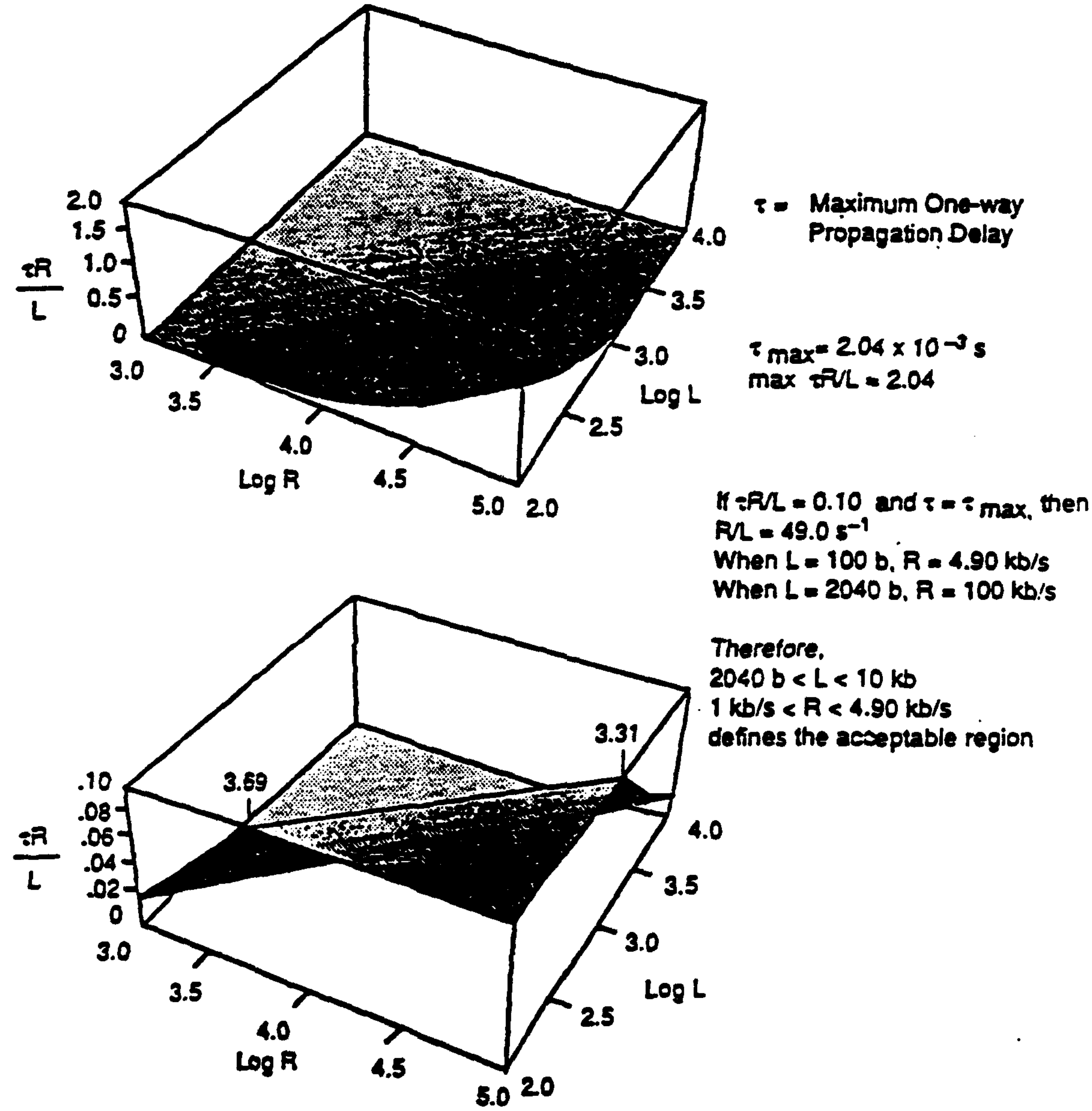
Although it is still desirable to minimize channel access delay with AOC type data links, access delay is relatively less important than higher throughput. Thus, we shall assume 1s for the nominal AOC near-real time access delay requirement. Note that this is five times longer than our assumption for an ATS-type link.

Delay will be a key parameter in the tradeoffs section. We shall refer to 200 ms, 1s, and 5s as our "real-time", "near-real-time", and "non-real time" channel access delay goals in comparing the delay performance of the various RA schemes.

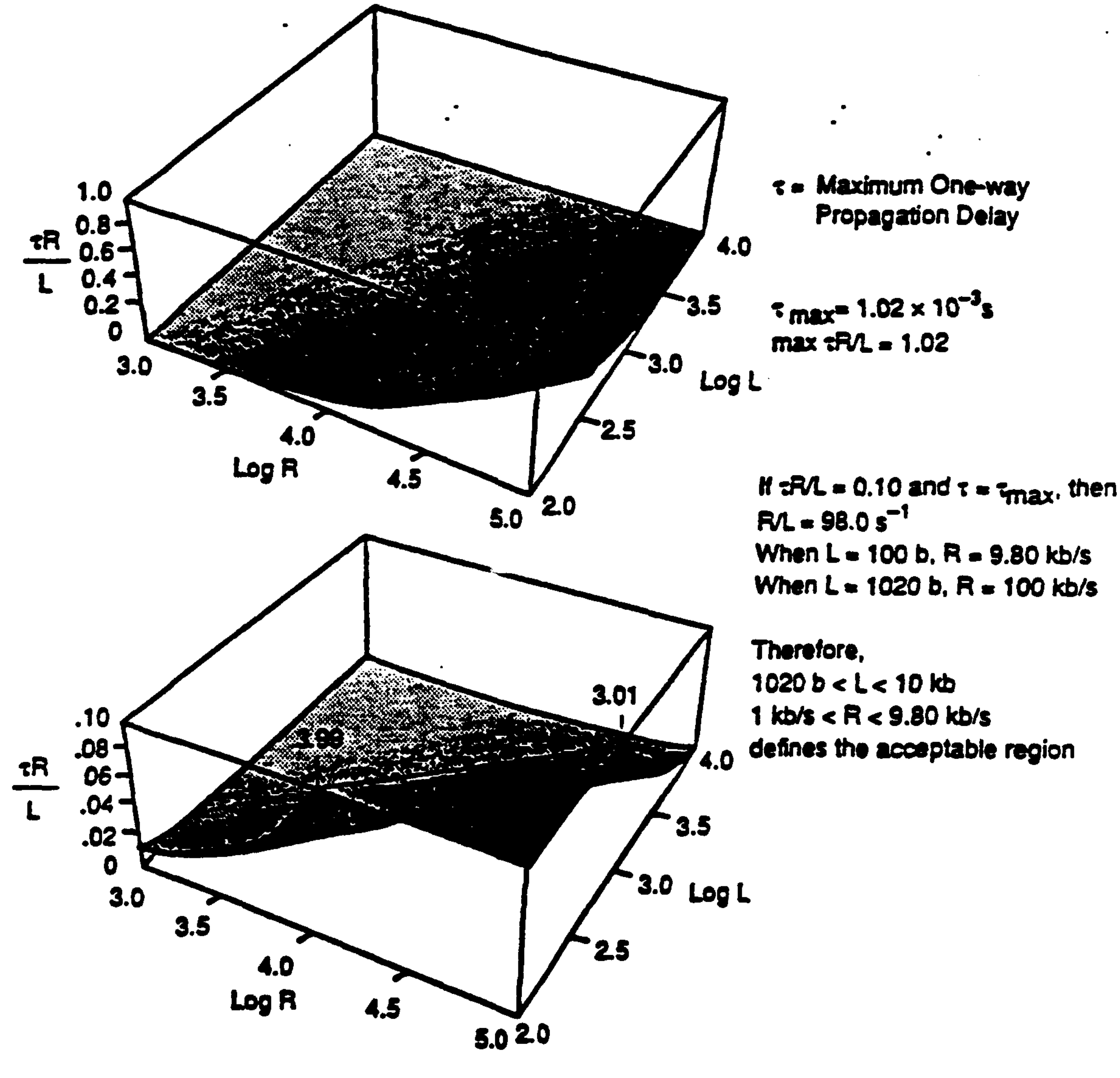
A Carrier Sense Multiple Access (CSMA) packet communications technique is used in ACARS and in the more advanced VDR which is to be consistent with the planned upgraded data communications standard called Aviation VHF Packet Communications (AVPAC). In CSMA a transceiver only transmits a message packet if the channel appears to be idle. For a given packet length (number of information bits in a message unit on the communications channel), as we have seen, the channel data rate governs the duration of the packet within the communications channel. The higher the data rate, the shorter the packet duration, and the less likely it is for that message to collide with another packet transmitted on the same channel. Collisions are always possible because more than one source transceiver may sense that the channel is idle and try to send a packet in the same epoch.

The maximum one-way propagation delay, τ, across the group of users is also an important parameter in CSMA schemes [8]. This is typically the duration one uses in sensing whether there is channel activity. Generally speaking, the throughput and delay performance of CSMA is improved over other RA schemes when the ratio of propagation delay to message duration $a = \tau/T$, is much less than unity. If a is too large, there is a point at which performance degrades by waiting too long for an idle channel. This will be shown quantitatively in the tradeoffs section of this paper.

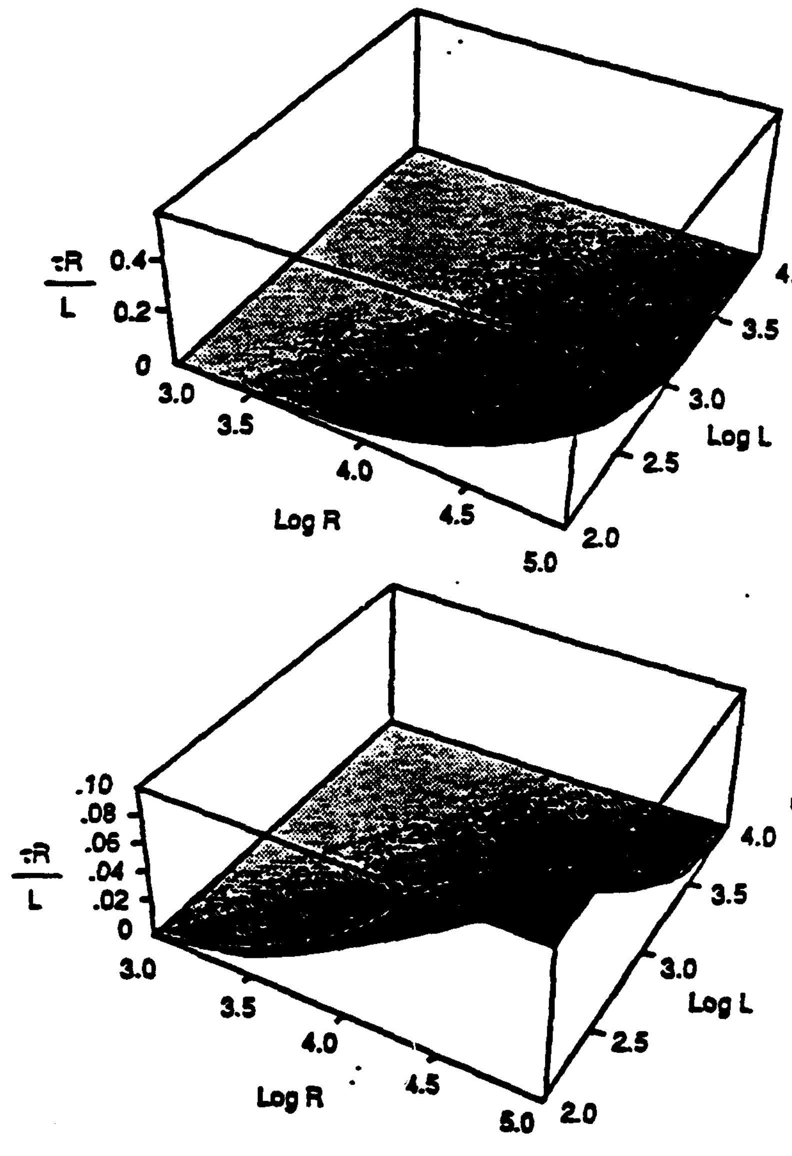
The minimum value of packet duration, $T = L/R$, to be expected in an ATC environment can be calculated assuming the minimum length message, say, $L_{min} = 200$ b, and maximum information rate, say, $R_{max} = 40$ kb/s. This yields a minimum packet duration of about $T_{min} = 5$ ms. Propagation delay to the radio horizon is greatest for an aircraft at the highest altitude, say, 72,000 ft. Using the well known formula, $r = \sqrt{2h}$, where h is the aircraft altitude expressed in feet, and r is the radio range in statute miles, one obtains a maximum propagation delay of $\tau_{max} = r_{max}/c = 2.04$ ms, where c is the speed of light. Thus, we see that typically, $a \leq 0.41 < 1$. In air-air communications this number can double. However, for the vast majority of ATC scenarios, a will be much smaller. For example, with 1000 b messages, an information rate of 20 kb/s, and an altitude of 18,000 ft, $a \leq 0.02 << 1$. Figures 11-1a, 11-1b, and 11-1c show the "acceptable" regions of L and R to achieve an a value of no more than 0.1; this assures the potential applicability of CSMA techniques.



Figures 11-1a. "Acceptable" Region for Packet Length (L) and Information Rate (R); Altitude = 72,000 ft



Figures 11-1b. "Acceptable" Region for Packet Length (L) and Information Rate (R); Altitude = 18,000 ft



τ = Maximum One-way Propagation Delay

τ_{max} = 0.51 x 10^{-3} s
max $\tau R/L$ = 0.51

If $\tau R/L$ = 0.10 and $\tau = \tau_{max}$, then
R/L = 196 s^{-1}
When L = 100 b, R = 19.6 kb/s
When L = 510 b, R = 100 kb/s

Therefore,
510 b < L < 10 kb
1 kb/s < R < 19.6 kb/s
defines the acceptable region

Figures 11-1c. "Acceptable" Region for Packet Length (L) and Information Rate (R);
Altitude = 4,500 ft

For reference later the maximum packet duration will be taken as 0.5 s, obtained by assuming that the maximum packet length is L_{max} = 10,000 b and that the minimum information rate is R_{min} = 20,000 b/s.

In normal operations in today's ATC environment, a group of pilots under the control of one controller to which they are assigned use a simple form of CSMA. Each pilot of the group typically listens to the common channel frequency and if the channel is idle, activates a push-to-talk (PTT) button and talks if there is a need to speak. The pilots follow an understood discipline of not interrupting each other (listen before talk (LBT)) except in unusual circumstances such as an emergency. If two pilots begin speaking at the same time, one typically backs off until the other is finished. In this case, the protocol is actually more sophisticated than the basic CSMA since a collision detection (CD) feature is inherent. In general, the performance of RA schemes can be improved by adding a CD capability to free the channel more quickly if overlapping messages are sensed. Naturally, there is a premium on sensing a collision early. The maximum number of aircraft that can be safely handled by a single controller is, of course, scenario dependent. However, in normal busy circumstances, a number of ten pilots to one controller is not uncommon. Much of a pilot's speech is in response to and acknowledgment of the controller's instructions.

Note that CSMA schemes with or without CD require an ability to receive transmissions from all users in the group. If all users are utilizing the same transmit frequency, as is usually the case, the channel is characterized as a broadcast channel, i.e., each member of the group can "hear" any other member.

Also observe that a user not able to monitor its own transmission cannot perform CD, at least not immediately. Such a user, e.g., one that has only a half-duplex transceiver, i.e., one that can transmit or receive but not at the same time, must wait for feedback, or lack thereof, to determine whether a transmission was successfully received by the intended receiver. Unless a transmission is relayed on a different frequency, such as via a satellite, or at a different time, such as by a store and forward mechanism on another platform, the original transceiver cannot monitor its own transmission. Only users with full-duplex transceivers, i.e., that can receive and transmit at the same time, can perform immediate CD.

If there is a significant but limited delay in determining that a collision has occurred, it may be advantageous for each user to autonomously perform a collision resolution algorithm (CRA) as a means of improving channel throughput. This is possible on a satellite downlink broadcast channel as will be illustrated in the issues section.

Throughput can also be improved if the group of users and their controller are synchronized to common time with a degree of accuracy that is much less than the duration of a communications burst. In the case of Time Division Multiple Access (TDMA), for example, this corresponds to being able to fit the burst into a time slot without any significant guard time.

In the case of packet mode communication a common time accuracy effectively assures that the integrity of the entire packet is not nullified by a collision with another packet that only partially overlaps in time. Packets are usually equal length and are transmitted in time slots equal to the packet duration, T. Without this synchronism the vulnerability interval, i.e., the range of time when a collision with another single packet could occur, of a packet would be twice the packet duration, 2T, instead of just T, as with slotted transmission.

The question remains as to what degree of timing accuracy is reasonable to achieve this communications synchronism and throughput efficiency. Some "higher end" VHF airborne radios are able to achieve an absolute time accuracy in the order of 1 to 10 ms through various means including on-board navigation equipment, the use of a microprocessor, and input/output interfaces such as the standard buses described in ARINC-429 and MIL-STD-1553. The VDR is designed to correct its timing accuracy of a millisecond or so every few minutes by receiving timing information from more accurate ground based radio clocks. The less complex/expensive radios found in the general aviation (GA) category have no provision for absolute timing. Also, typical (GA) radios possess a frequency stability ranging from only 30 to 60 parts per million (ppm).

It is assumed that any radio able to obtain timing inputs is able to do so sufficiently often that the frequency stability question is moot. Thus, the best uncorrected absolute timing accuracy of current airborne radio equipment is no better than about 1 ms. In addition propagation delay differences between an aircraft and its ground control can vary from essentially zero to the product of the radio range $r = \sqrt{2h}$ to the horizon in statute miles and the number of seconds it takes the signal to propagate one statute mile, i.e., $\sqrt{2h} \times 5.36\ \mu s$, where h is

the aircraft altitude in feet. This product is approximately 0.5 ms, 1 ms, and 2 ms for altitudes of h = 4500 ft, 18,000 ft, and 72,000 ft, for example, cf., figure 11-1. This adds to the inherent timing uncertainty of radios with a time keeping capability.

It follows that for radios with the usual time sources, without further refinement of timing accuracy, the communications burst in a TDMA time slot, or packet duration, as the case may be, must be much greater than several ms, i.e., several tens of ms. Otherwise, throughput efficiency is lost due to excessive guard time to protect against burst or packet collisions. Alternatively, if the burst or packet lengths are as short as 5 or 10 ms, an improved timing system must be introduced to achieve slotted throughput efficiencies. Of course, radios without any timing sources also would need a good timing system to gain from slotted techniques.

In a new digital radio system as envisioned for the VDR, the airborne radios can key on the uplink transmissions from the ground. The ground radio can maintain accurate time and serve as a master timing station transmitting bursts only at the beginning of a slot. Except for the need to correct for propagation delay differences, the derived airborne radio accuracy is then approximately equal to the reciprocal of the radio's IF bandwidth which in our cases of interest is typically tens of kHz, i.e., the uncertainty is tens of microseconds. However, with propagation delay uncertainties of one or two ms, the previous conclusion is basically unchanged. A more accurate ranging system is necessary to achieve the efficiency benefits of a slotted access scheme.

11.3 ISSUES

We begin this section by defining various RA schemes of potential interest and presenting information on their theoretical throughput/delay performance and their degree of stability under heavy channel loading. Various ALOHA schemes and their modifications will be treated first. Then basic collision resolution (tree) algorithms (CRAs) will be discussed. This will be followed by several CSMA schemes including the protocols previously considered and currently favored for the VDR development by the AEEC design subcommittee, namely, virtual-time CSMA and non-adaptive p-persistent CSMA, respectively.

It is noted that well-known Frequency, Time, and Code Division Multiple Access (FDMA, TDMA, and CDMA) techniques can be combined with the notion of RA in the following sense.

In FDMA, users could contend for a set of allocated frequencies by each randomly selecting one of these frequencies for transmission, and living with the possibility that at least one other user has selected that frequency in the same epoch. Similarly, users could content for time slots by picking one at random, again with the possibility of interference from other users choosing the same slot. Finally, in CDMA users could pick among the permissible codes at random with some likelihood that more than one user would select the same code. The intrasystem interference mechanisms particular to these hybrid access modes, other than the contention for the same frequency, time slot, or code resource, respectively, are the same as those associated with fixed (and demand) assigned FDMA, TDMA, and CDMA.

The contention aspect, which is most clearly viewed as occurring in the time domain on a single frequency broadcast channel employed by all users, is inherent to the ALOHA, CRA and CSMA protocols to be discussed. Hence, the FDMA, TDMA and CDMA hybrids mentioned above will not be treated any further in this memo.

11.3.1 Pure and Slotted ALOHA

The fundamental difference between Pure and Slotted ALOHA is that in Slotted ALOHA, users have accurate time information so as to transmit packets in well-defined time slots just equal to the packet duration. Thus, when collisions occur in Slotted ALOHA the packets completely overlap in time. If there is any overlap between two packets it is usually assumed in the theoretical models that neither packet gets through, i.e., neither packet is received correctly by the destination receiver.

Let G be the average number of new and retransmitted (old) packets per packet interval, T. Let S be the expected throughput, i.e., average number of packets per packet interval that are received correctly (do not collide). Then it is possible to express the throughputs for Pure and Slotted ALOHA very simply:

$$S = G e^{-2G}, \text{ Pure ALOHA} \quad (1a)$$

$$S = G e^{-G}, \text{ Slotted ALOHA.} \quad (1b)$$

As shown in figure 11-2 [9, fig. 3, p. 473; 10], the maximum throughput of 18% and 37% (1/2e and 1/e) for Pure and Slotted ALOHA occurs for G = 1/2 and G = 1, respectively. Although Slotted ALOHA has better throughput, this protocol requires absolute timing throughout the system to maintain the integrity of the slot structure. Note that other (CSMA) schemes to be discussed later have much better throughput than Slotted ALOHA.

In the Pure and Slotted ALOHA algorithm a new packet is transmitted in the very next packet interval, and an old packet is usually retransmitted in one of the next K intervals of P packet slots each. K is a constant large enough to assure a low probability of re-collision if the packets transmitted by a few users have just collided in the preceding epoch. The particular slot interval of the 1 to K consecutive intervals of the present epoch chosen for transmission by a given user is selected by a pseudorandom number generator. A user knows whether a given packet was successfully transmitted if a positive acknowledgment is returned from the intended receiver after a delay of d packet slots. If a negative acknowledgment is returned or if no acknowledgment was returned after a predetermined "time-out" period, the user knows that the packet transmission was unsuccessful, and a retransmission is attempted. For simplicity it is assumed that the time-out period is identical to the round-trip delay, d.

The expected packet delay for each user is computed as follows. The delay in waiting for each selected retransmission interval of duration P to arrive is (K-1) P/2 packet slots. Let $q = S/G$ be the probability of successful packet transmission, and let i be the number of retransmissions of a given packet before it is successfully received. This occurs with probability $(1-q)^i q$, $i = 0, 1, 2, \ldots$, since the successive transmission trials are assumed to be independent. Inserting the round-trip delay d for each trial, and the maximum propagation delay plus transmission delay, a + 1, of the last successful packet transmission, we have

$$\begin{aligned} D &= \sum_{i=0}^{\infty}\left[a + 1 + \left(d + \frac{(K-1)P}{2}\right)i\right](1-q)^i q \\ &= a + 1 + \left(d + \frac{(K-1)P}{2}\right)\frac{1-q}{q} = \left(d + \frac{(K-1)P}{2}\right)\left(\frac{G}{S} - 1\right) + a + 1 \end{aligned} \qquad (2)$$

Typical delay performance of these two ALOHA schemes is shown in figure 11-3. See the text associated with equations (7), (9) and (11) of the Collision Resolution Algorithms section for a rationale for the value d = 3.

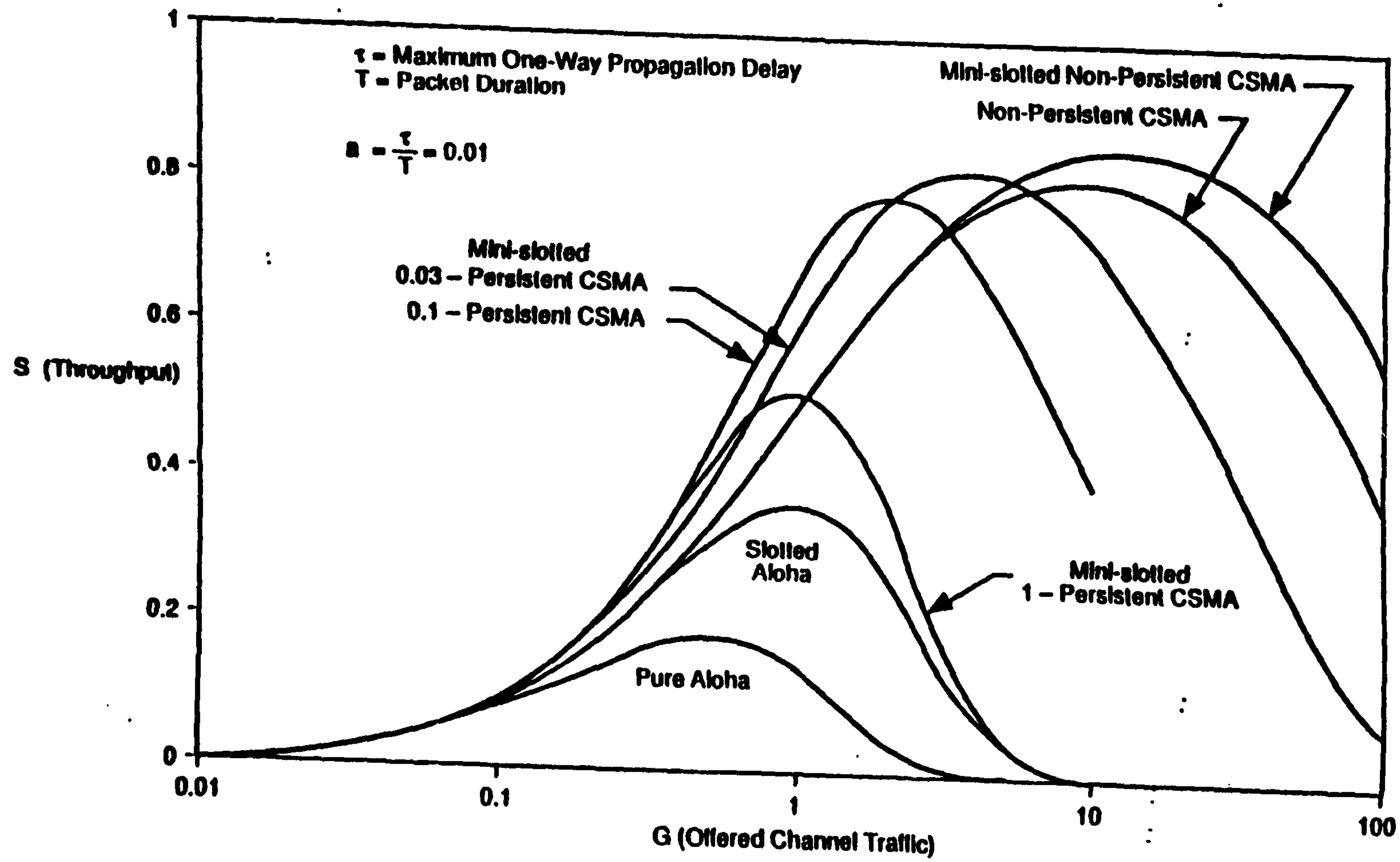


Figure 11-2. Theoretical Throughput vs. Offered Channel Traffic

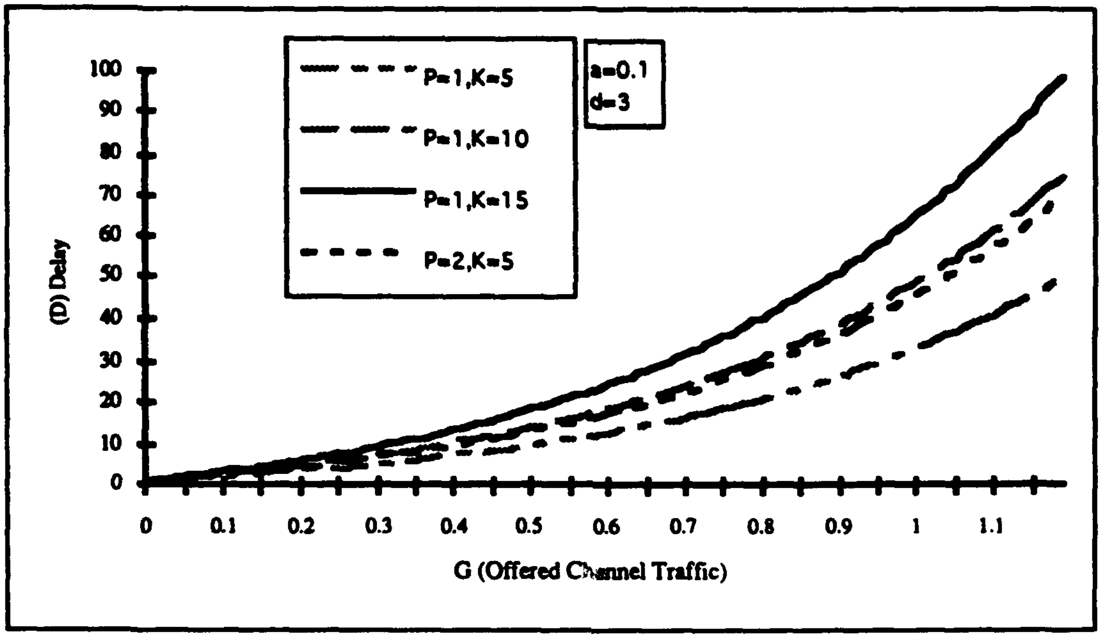


Figure 11-3a. Delay vs. Channel Loading for Pure ALOHA

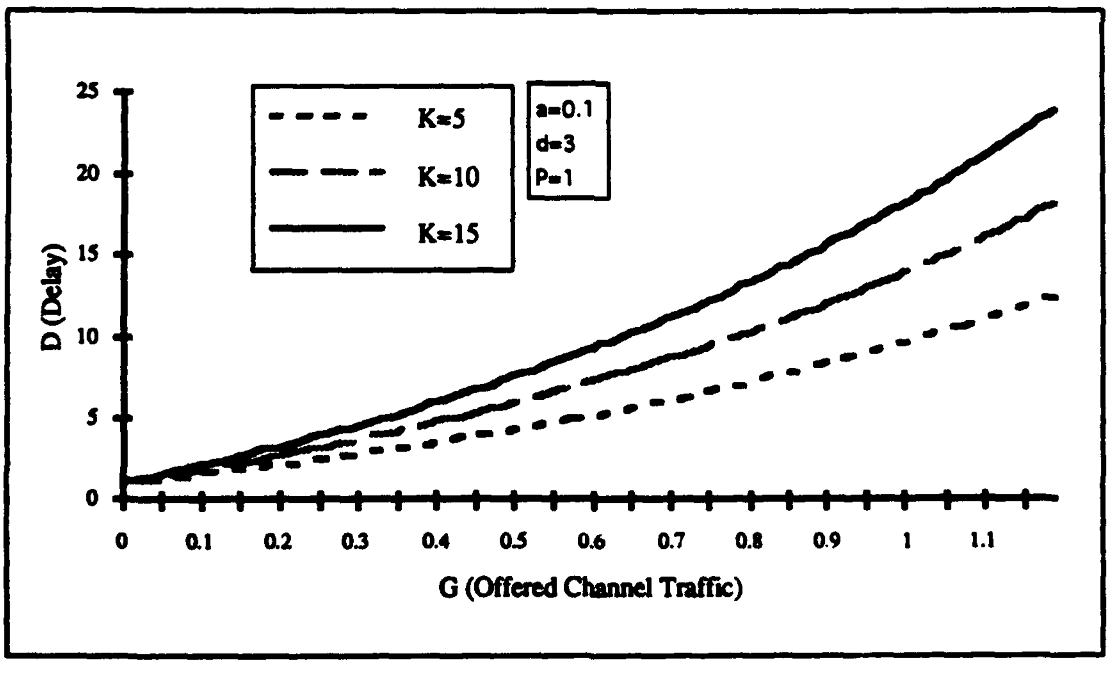


Figure 11-3b. Delay vs. Channel Loading for Slotted ALOHA

From equations (1) and (2), note that the delay D is unbounded in G, and that the delay is a two-valued function of the throughput S. Also notice that the throughput rapidly declines as the channel traffic increases beyond the point of maximum throughput. Thus, Pure and Slotted ALOHA are unstable algorithms and should be used with caution. It would be prudent to attempt to operate with a channel loading well below that required for maximum throughput, e.g., G = 0.18 and G = 0.36, for S = 12.5% and S = 25%, respectively, cf., equation (1). Unfortunately, this would require close monitoring of the channel by a central controller as well as a mechanism to reduce channel loading when the operating point is exceeded by any significant amount.

Modified ALOHA is a more practical scheme to accomplish stability because the "braking" mechanism does not depend on a centralized controller; each user acts autonomously. When experiencing a packet collision, instead of transmitting in the very next interval, as in Pure or Slotted ALOHA, a user retransmits the packet in an interval selected at random from the next $2^i \times K$ intervals, where i is the number of collisions that have already occurred with this packet, and K is an adjustable constant like ten or fifteen. That is, if collisions persist, the interval where the packet will be retransmitted continues to double; since each user acts independently and at random, the contending packets are much less likely to collide with each retransmission. When a packet is successfully transmitted (without a collision) by a user, the next epoch reverts to K intervals and the process repeats. Simulations have shown that this greatly stabilizes the ALOHA schemes without much increase in expected delay.

11.3.2 Reservation ALOHA

This usually is a hybrid scheme where a RA Pure or Slotted ALOHA technique is used on a "request" subchannel in attempting to reserve a subsequent frequency or time interval for dedicated transmission. As one would expect, and as shown in figure 11-4 [9, 11] for Slotted ALOHA, better throughput and delay is obtained at the higher throughput values, compared to Pure and Slotted ALOHA, at the expense of the dedicated channel resources.

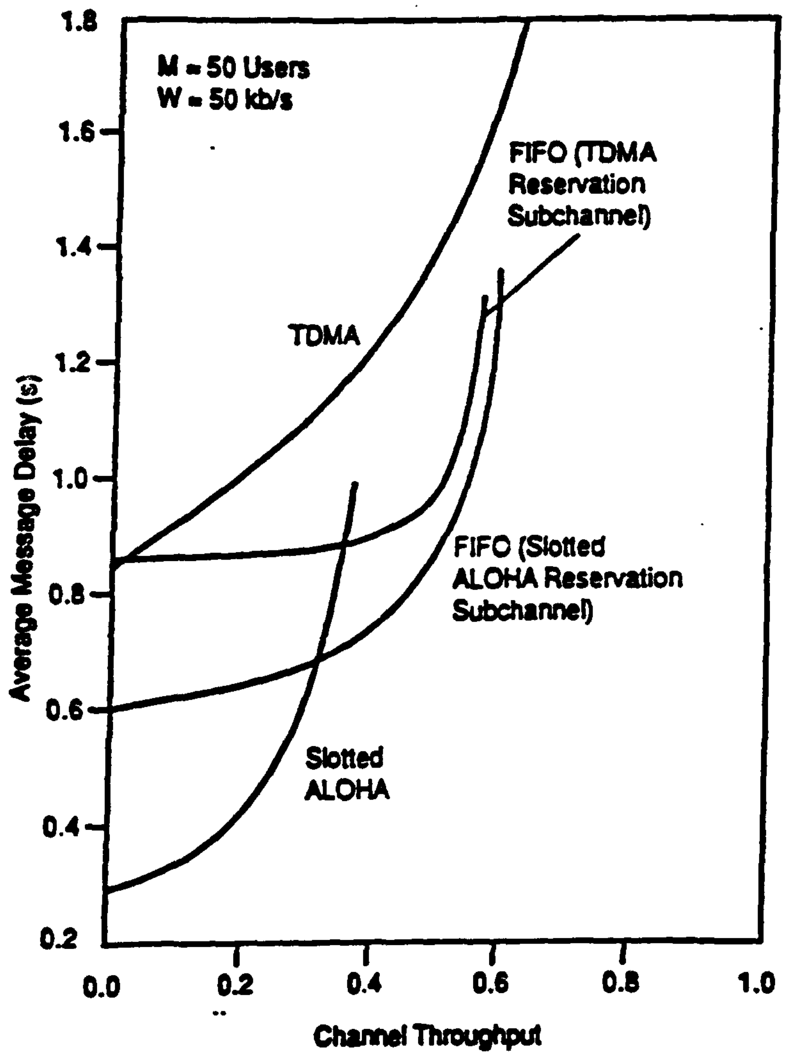


Figure 11-4. Slotted ALOHA, TDMA, and First-In, First-Out (FIFO) Reservation: Delay Throughput Tradeoff for 50 Users and Single-Packet Messages in a Satellite Environment [11]

11.3.3 Collision Resolution Algorithms

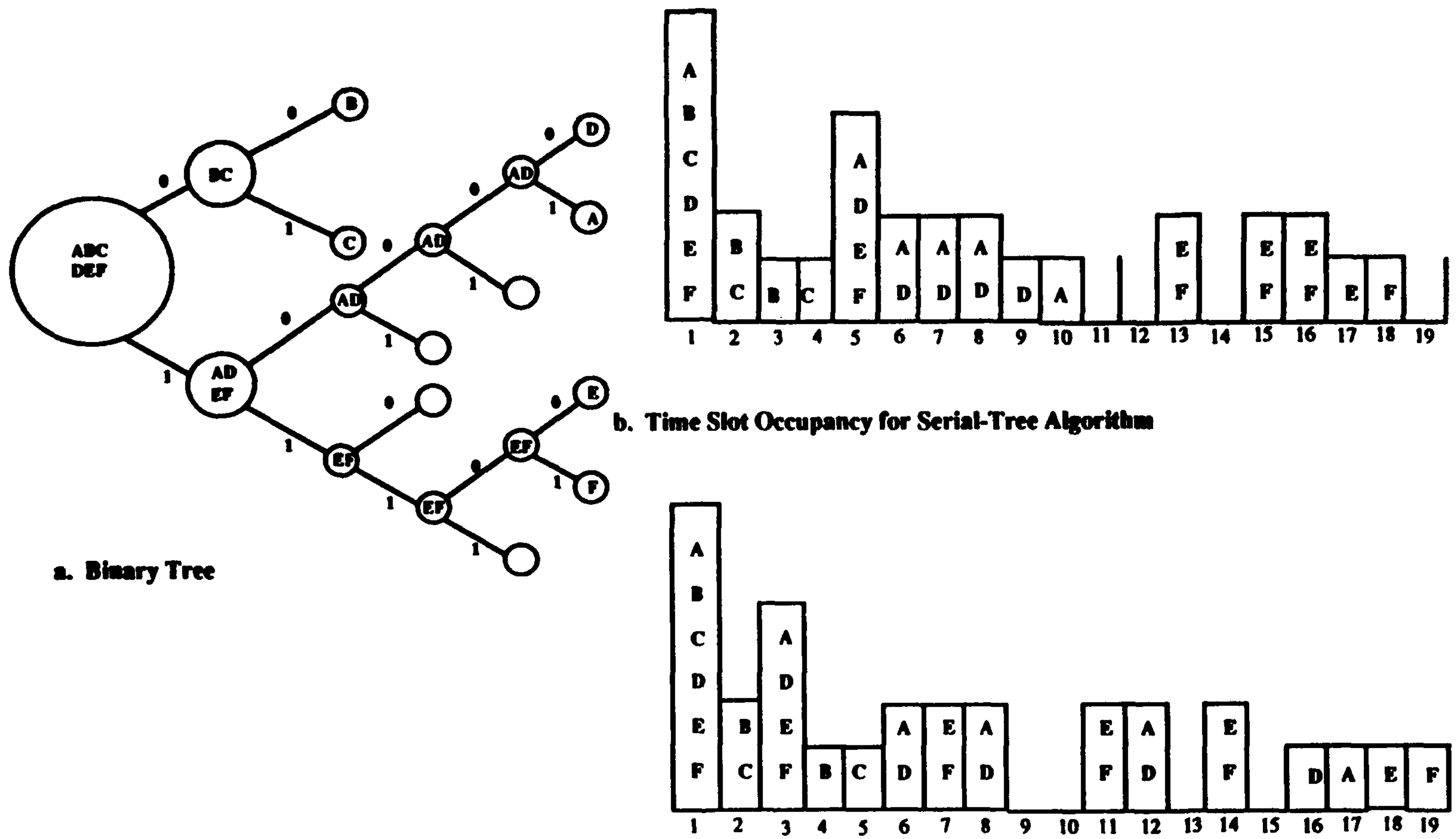
Collision Resolution Algorithms (CRAs) are another class of RA schemes that can be applied to groups of low duty factor users attempting to gain access to a communications channel with limited capacity. As with reservation ALOHA, CRAs could also be used on a request channel for accomplishing the demand assignment of other channels to requesting users. CRAs operate with a well defined time slot structure as with Slotted ALOHA.

The so-called "inference seeking" CRA algorithms have the major advantage of stability over Slotted ALOHA and comparable throughput and delay performance. This is accomplished by autonomous users making logical inferences based on whether collisions occurred in previous slots. By observing the past history of the broadcast channel and operating in a way about to be described, each given user among a group that has "collided" can be assured of successful packet transmission by the end of the "conflict resolution phase" whose length is proportional to the number of users that collided. CRAs are also called "tree" algorithms since the logical inference followed by each user can be described as a binary tree structure.

Basic Algorithm

The basic serial-tree algorithm is best explained by way of example. Suppose there are N=6 users designated A, B, C, D, E and F which have collided or created mutual interference on the channel by transmitting their packets in the same packet interval. This situation is depicted in figure 11-5 by the entry in the root node of the binary tree and the occupancy of the first time slot. A fundamental assumption required in this RA algorithm is that each user can determine whether there is a packet collision on the channel in any given time slot.

The general rule is as follows. At the beginning of every collision resolution phase (including each smaller resolution phase that is part of a larger one), each user involved in the collision flips a 0 - 1 coin. Each user flipping 0 transmits in the next slot. Each user flipping 1 does not transmit until the number of collision-free slots exceeds the number of collision slots by one during the current resolution phase; these users then transmit in the very next slot. Note that at no time does any user know the exact number of users colliding.



a. Binary Tree

b. Time Slot Occupancy for Serial-Tree Algorithm

c. Time Slot Occupancy for Parallel-Tree Algorithm

Figure 11-5. Collision Resolution Example Using the Basic Algorithm

Furthermore, it is unnecessary for any user to distinguish a blank slot from a successful single packet transmission.

The parallel-tree algorithm is similar to the serial-tree algorithm except that the users with colliding packets go the end of the waiting line instead of remaining at the front. Thus, the serial-tree time slot occupancy of figure 11-5b is transformed into that of 11-5c. This corresponds to traversing the binary tree of figure 11-5a in parallel rather than serially.

Maximum Average Stable Throughput

In order to describe the throughput and delay performance of the RA system discussed, it is convenient to define a number of random variables and deal with their expected or average values from a statistical point of view. The following definitions will be employed:

- X = number of packets that have collided in the initial slot of a collision resolution interval (CRI)
- Y = length of this CRI in packet slots
- L_N = $E(Y|X = N)$ average length in packet slots of a CRI given that N packets collided initially
- D' = average number of packet slots between the arrival of a packet at a given transmitter and the onset of that packet's successful transmission
- S = average total rate of new packet arrivals in the network in packets per packet slot, i.e., throughput.

A RA scheme is said to be stable if D' is finite. Massey showed that this condition is equivalent to $E(Y)$ being finite for the basic algorithms [12]. Also, Massey showed that

$$L_N = \alpha N - 1 \tag{3}$$

is a very good approximation, where

$$\alpha = 2.89, \quad N \geq 4. \tag{4}$$

The average number of new packet arrivals during a certain CRI must not exceed the number of successfully transmitted packets during the same interval or the scheme will be unstable, i.e.,

$$SL_N \leq N \tag{5}$$

must hold for stability. Using (3) and (4), this implies that

$$S < 1/\alpha = 0.346. \tag{6}$$

Note that such an upper bound on S is equivalent to the maximum average throughput of a stable RA scheme. For reference, the corresponding value for Slotted ALOHA, which is unstable, is $1/e = 0.368$.

Expected Packet Delay

Bounds on the average packet delay D′ assuming zero propagation and feedback channel delays are shown in figure 11-6. Although the maximum average stable throughput is 0.346 packets/slot, shorter delays are obtained if one can operate in the vicinity of $S = 0.25$, i.e., on the knee of the delay-throughput curve. As noted earlier for Pure and Slotted ALOHA, maintaining this operating point is a non-trivial matter. A realistic analysis of delay for the serial-tree algorithm and a VHF A/G channel follows.

Let the worst-case round-trip propagation delay in packet slots between any aircraft user and the ground system be D_r. Let any processing and transmission time in packet slots of a transceiver at the ground system site used for the purpose of returning a feedback message be d_p. This processing could be for the purpose of generating a positive or negative acknowledgment corresponding to a successful single packet transmission or a collision, i.e., an ACK or NAK, respectively. A LAK could designate a no return message condition indicating that nothing was perceived as being transmitted on the aircraft to ground channel. With these definitions the result of transmissions in packet slot i cannot be utilized by the most distant transmitter earlier than slot i + d, where

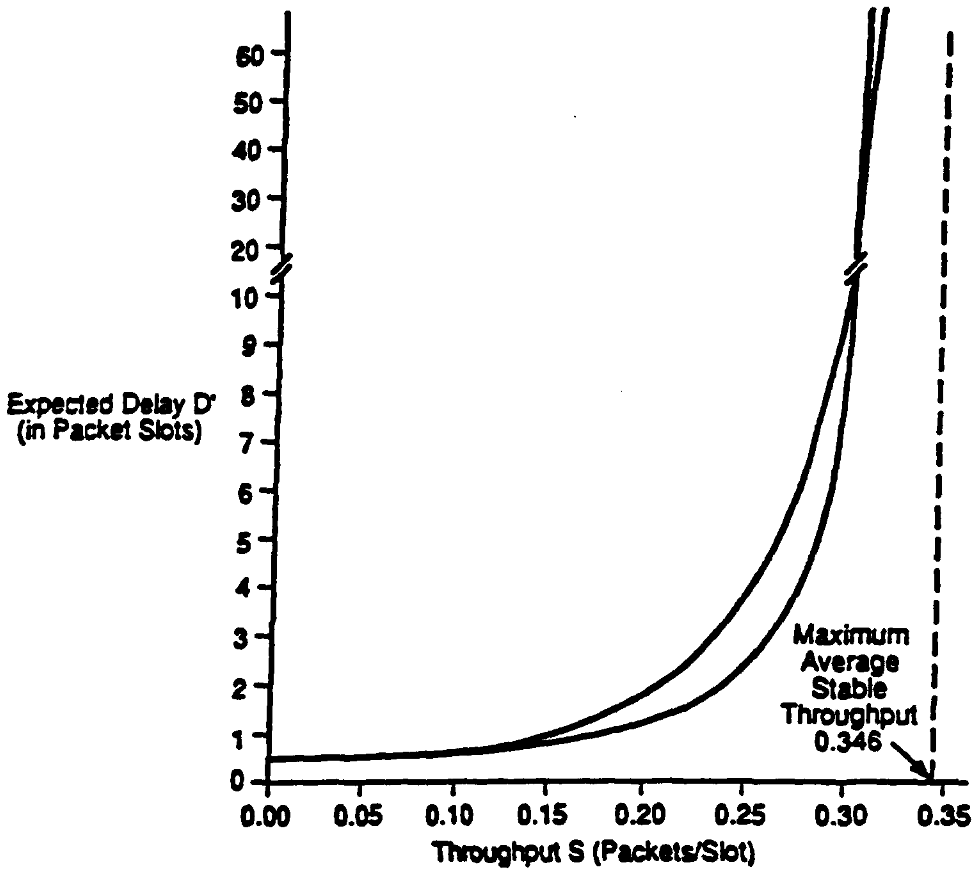


Figure 11-6. Bounds on Serial-Tree Delay [12]

$$d = \mathrm{INT}\,(D_r + d_p) + 1 \qquad (7)$$

and where INT (x) denotes the least integer no less than x.

As Massey suggested, the simplest way to extend the RA algorithm to account for this delay of d slots is to partition the original channel into d interleaved subchannels, where each such subchannel, in effect, has no propagation or processing delay. Thus, the jth subchannel consists of slots j, $j + d$, $j + 2d$, etc., of the original channel. The RA algorithm can be executed independently on each subchannel without any loss of generality.

With the basic algorithm a packet arriving in slot i is assigned to the (i + 1st) subchannel. Each subchannel still sees an average packet arrival rate of S packets/slot. On the average, a new packet must wait 1/2 slot before entering the new packet list for its corresponding subchannel. Once in the subchannel the new packet must wait an additional $D' - 1/2$ slots on the average in that subchannel before it can be transmitted, where D' is the expected delay of the original channel with $d = 1$. Hence, the expected packet delay D for the basic algorithm using this interleaved approach with propagation and feedback delays is

$$D = d(D' - 1/2) - 1/2. \qquad (8)$$

The maximum round-trip propagation delay D_r is given by twice the maximum propagation delay τ between an aircraft and the ground site divided by the packet or slot duration T:

$$D_r = 2\tau/T = 2a. \qquad (9)$$

As we have seen in the background section of this paper

$$0 < D_r < 1. \qquad (10)$$

The processing time in slots may include the calculation of ACK/NAK packet bits and the transmission (or reception) time of the feedback packet. It is assumed that the time for feedback packet calculations is negligible compared to its transmission time, however. For

simplicity let the feedback packet length be the same as the regular packet length. Hence, the processing delay d_p in slots is taken as

$$d_p = 1. \tag{11}$$

This, the sum of the roundtrip and processing delays implies that d is 3, from equation (7). From equation (8) this means an expected packet delay of D = 8 slots if the operating point is at D′ = 3, on the knee of the delay curve of figure 11-6. This delay would correspond to the average waiting time in gaining access to a dedicated communications channel in a demand assignment system, for example.

If the real-time (near-real time) channel access delay is not to exceed 200 ms (1s), as postulated in the background section then the slot duration should not exceed 200 ms/8 = 25 ms (1s/8 = 125 ms). Because of the statistical variation, the slots should probably be significantly shorter than that.

Collision Detection Model

Massey's work on introducing channel errors or collision detection into the basic CRA algorithm has been extended to include both aspects simultaneously [13]. Let β be the probability that no packet is transmitted over the channel but the user perceives that a packet collision has occurred. Let ε be the probability that a single packet is transmitted but that a collision is perceived. As in Massey's channel error model the probabilities of perceiving no packet when one was transmitted and perceiving a single packet when none was transmitted are taken as zero; finally, a collision is perceived with probability one if a collision indeed occurs.

The basic algorithm model is now perturbed in another way to accommodate the possibility of sensing blanks, collisions, or single packet transmissions on the channel. If a blank or a collision is sensed, each user shortens the current slot by an amount determined by the sensing time and immediately begins the next slot. Note that this can complicate the maintenance of slot synchronization but can only increase the maximum stable throughput and shorten the expected delay.

Let b and c be the fraction of a slot required to sense a blank and a collision, respectively. Typically, $b < c$ since a blank involves relatively small received power unless there is a significant amount of ambient interference. A blank is sensed if one or more carrier signals are not sensed.

It should be easy to distinguish a single packet from a collision since a special bit pattern can be used to identify or authenticate an uncorrupted packet. This pattern should be placed as early as possible in the packet header to shorten c, since a collision will be recognized as soon as at least one carrier is sensed but the special bit pattern does not appear at the demodulator output by the allotted time.

It is roughly estimated that a collision can be sensed within the first 100 b of the packet. For the range of the packet lengths estimated in the background section, this suggests that

$$0.01 \le c \le 0.5 \tag{12}$$

is a reasonable approximation. Single packet slots will be unchanged in a CRA with carrier sensing. However, blank and collision slots will be shortened by $1 - b$ and $1 - c$, respectively.

With these definitions it can be shown (see [13]) that the expected length of the CRI with collision detection and channel errors is

$$L_N = E(Y|X = N, \text{collision detection, channel errors}) \lessapprox \tag{13}$$

$$\left(1 + \frac{(0.445 + 0.555\varepsilon)(1+\beta)b + [1.445 - 0.445\varepsilon - (2.445 - 1.445\varepsilon)\beta]c}{(1-\varepsilon)(1-2\beta)}\right)N - c, \quad N \ge 4$$

The basic algorithm is stable if S is less than the inverse of the coefficient of N in equation (13). (See figure 11-7.) As a simplified limiting example of how collision detection can improve the expected stable throughput without any channel errors, suppose $\beta = \varepsilon = 0$, and $b = c$. Then the coefficient of N reduces to $1 + 1.89\,c$ and

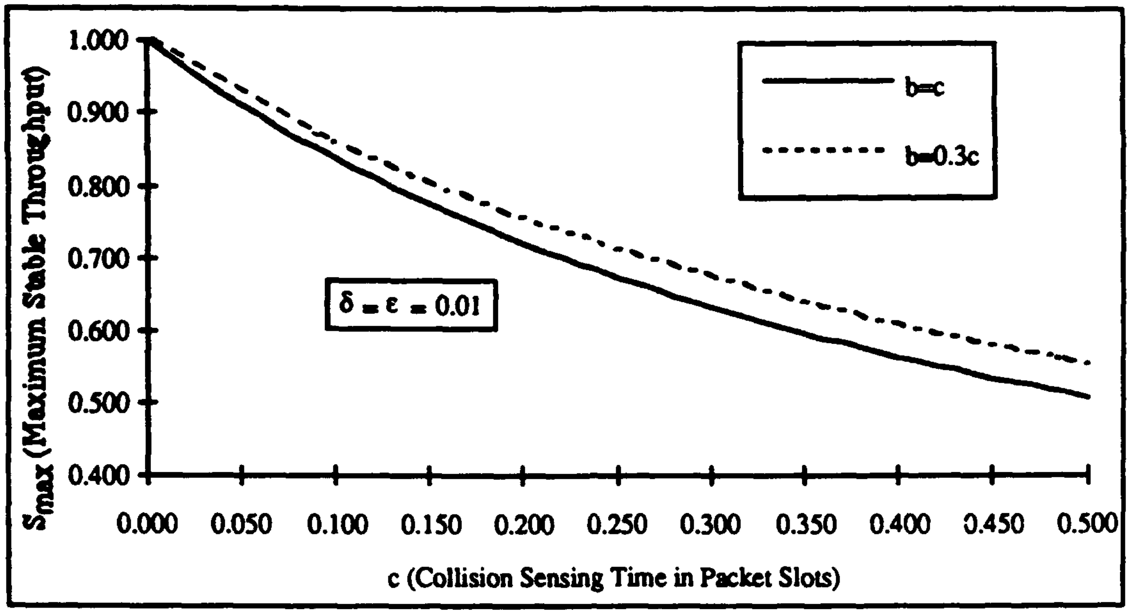


Figure 11-7a. Throughput vs. Collision Sensing Response
(S < Inverse of Coefficient of N in Equation (13))($\varepsilon = 0.01$)

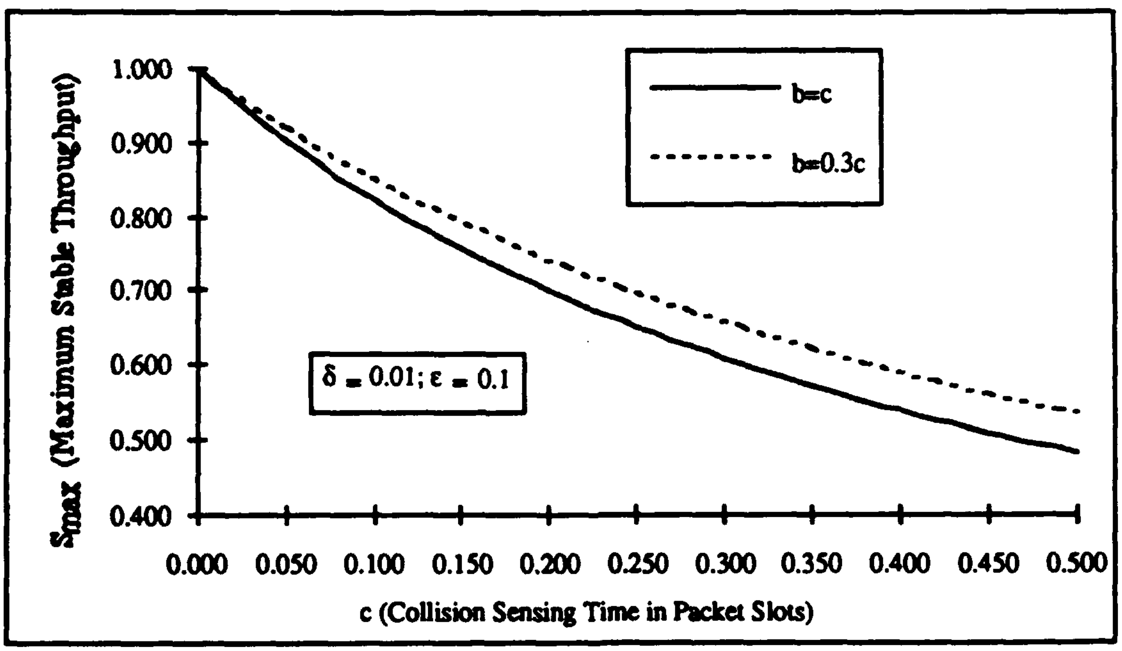


Figure 11-7b. Throughput vs. Collision Sensing Response
(S < Inverse of Coefficient of N in Equation (13))($\varepsilon = 0.1$)

$$S < \frac{1}{1 + 1.8\nu c} \tag{14}$$

for stability. Using equation (12) we see that the throughput, S could range between 0.51 and 0.98 for packets between 200 b and 10,000 b in length, respectively.

11.3.4 Mini-Slotted Alternating Priorities

In this section we analyze a typical algorithm from the class known as mini-slotted alternating priorities [9, p. 482, 14]. Consider a group of n users, each identified by a unique integer $i = 1, 2, ..., n$, which also serves as an indication of user priority. User i has priority over user j, where $i < j \leq n$. The priorities of these users can be shuffled periodically, either automatically according to some predetermined algorithm, or upon command by a central controller, perhaps in association with the recomposition or reassignment of the group of users. However, priority only matters upon "start-up"; in the steady state each user is treated equitably. In a sense this protocol is like a "token ring" where authority to transmit is passed around the group of users.

The broadcast channel is either idle, i.e., no user is transmitting, or it is occupied. If idle, channel time is partitioned into consecutive "mini-slots" each of duration w. If occupied, i.e., a user is transmitting, channel time is partitioned into consecutive packet slots each of duration T. All users must have relative time, accurate to a small fraction of a mini-slot.

If users have packets to send they do so in order of their assigned priority. A transmitting user i sends all its packets in consecutive packet slots and includes its priority index in each packet transmitted. All other users receive these packets and can therefore determine which user is transmitting and when that user is finished sending its packets. When the channel becomes idle the next user j in order of priority and with at least one packet to transmit will occupy the channel with all its packets. The number of consecutive mini-slots on the channel between the transmissions of user i and user $j > i$ is exactly $j - i$. User j knows it can send its packets when it observes that this many consecutive mini-slots have occurred, i.e., no other user with priority higher than user j but lower than user i has packets to transmit.

Imagine these mini-slots on a time scale relative to a fixed local observation point. The first mini-slot in the current series (in "steady state") of mini-slots begins at time t_0 with the end of transmission by the previous user, of index i, say, of the channel. This end of transmission has propagated to all users of the channel by time $t_0 + \tau$, where τ is the maximum one-way propagation delay across the group of n users. Suppose it takes each user a time δ to detect the presence or absence of a transmission on the channel. If the next user with a packet to transmit is user $j > i$, then the duration of the current idle channel state is between $(j - i)\,\delta$ and $(j - i)\,(\tau + \delta)$. In general, each node adds between δ and $\tau + \delta$ to the idle gap between transmissions. On a steady state statistical basis it is assumed that on the average each user adds

$$w = \delta + \tau/2 \tag{15}$$

to the idle gap; thus, w is taken as the mini-slot duration.

Assuming the usual Poisson model where each user generates packets at the same average rate of g packets per packet interval T, the total average channel loading of the group of n users is $G = ng$ packets per packet slot. Since there is no possibility of packet collision in this scheme, the throughput is $S = G$. However, if an average transmission of a given user with at least one packet to transmit is z packets, and if there are an average of s mini-slots between two different users' adjacent transmissions, then the channel utilization is

$$U = \frac{zT}{zT + sw} = \frac{z}{z + sv} \tag{16}$$

where $v = w/T = \frac{\delta}{T} + \frac{a}{2}$, from equation (15). We need to compute z and s.

A typical user must wait an average normalized "cycle" length of

$$Q = n(z + sv) \tag{17}$$

packet slots between two successive opportunities to transmit its accumulated packets. The probability that any user will not have generated a packet in this interval is, cf., equation (0) in the background section,

$$P_0(Q) = e^{-gQ}. \tag{18a}$$

The probability that any user generates <u>at least</u> one packet during this cycle is

$$P(Q) = \sum_{k=1}^{\infty} P_k(Q) = 1 - P_0(Q) = 1 - e^{-gQ}. \tag{18b}$$

From equation (0), and by definition, the expected number of packets arriving at any user's transmit buffer during one cycle is

$$E(k) = \sum_{k=0}^{\infty} kP_k(Q) = \sum_{k=1}^{\infty} kP_k(Q) = g\,Q. \tag{19}$$

If $l(k)$ is the average duration of the interval associated with an arbitrary user's transmission of the k consecutive packets that arrive in one cycle, we have

$$l(k) = kT + w, \;\; k = 0, 1, 2, \ldots \;\; . \tag{20a}$$

Normalizing by the packet duration, T, the expected length of the interval between successive users is

$$\begin{aligned} \frac{1}{T}\sum_{k=0}^{\infty} l(k)P_k(Q) &= \sum_{k=1}^{\infty} kP_k(Q) + v \\ &= E(k) + v = gQ + v \end{aligned} \tag{20b}$$

using equations (18b) and (19). Thus, we can substitute equation (20b) for the factor z + sv in equation (17):

$$Q = n(gQ + v) = GQ + nv. \tag{21a}$$

(This key step in the solution as well as the "cycle" approach to this problem is due to Dr. Warren J. Wilson.) Thus,

$$Q = \frac{nv}{1 - G} = \frac{n\left(\frac{\delta}{T} + \frac{a}{2}\right)}{1 - G},\ G = S < 1. \tag{21b}$$

Since Q is a queueing delay corresponding to the cycle time, we are more interested in the total delay, D, for a typical packet from the time it arrives at any user's buffer until it arrives at the destination receiver. Assuming a packet in the middle of a transmitted sequence, this delay can be expressed as, using equation (21b)

$$D = \frac{Q}{2} + 1 + a = \frac{nv/2}{1 - G} + 1 + a = \frac{n\left(\frac{\delta}{T} + \frac{a}{2}\right)}{2(1 - G)} + 1 + a. \tag{22}$$

Thus, this typical packet must wait about half the cycle time on the average; the 1 and a correspond to the packet's transmission and propagation delay, respectively. Delays for $n = 16$ and 32 with $a = 0.1$ and $\delta/T = 0.1$ are shown in figure 11-8 as a function of $G = S$.

Since there must be at least one (idle) mini-slot between every pair of adjacent transmissions, referring to equations (17) and (21a), note that $s > 1$ and $z < gQ$ must hold. The parameter z, the average value of k in the steady state, provided k is at least unity, is given by, cf., equations (18b) and (19)

$$\begin{aligned} z &= \sum_{k=1}^{\infty} k\,[P_k(Q) \,.\, P(Q)] = P(Q)\,E(k) \\ &= (1 - e^{-gQ})\,gQ = \left(1 - e^{-\frac{GQ}{n}}\right)\frac{GQ}{n}. \end{aligned} \tag{23a}$$

Now from equation (16)

$$s = \frac{1}{v}\left(\frac{Q}{n} - z\right) = \frac{\left(\frac{Q}{n} - z\right)}{\left(\frac{\delta}{T} + \frac{a}{2}\right)}\ . \tag{23b}$$

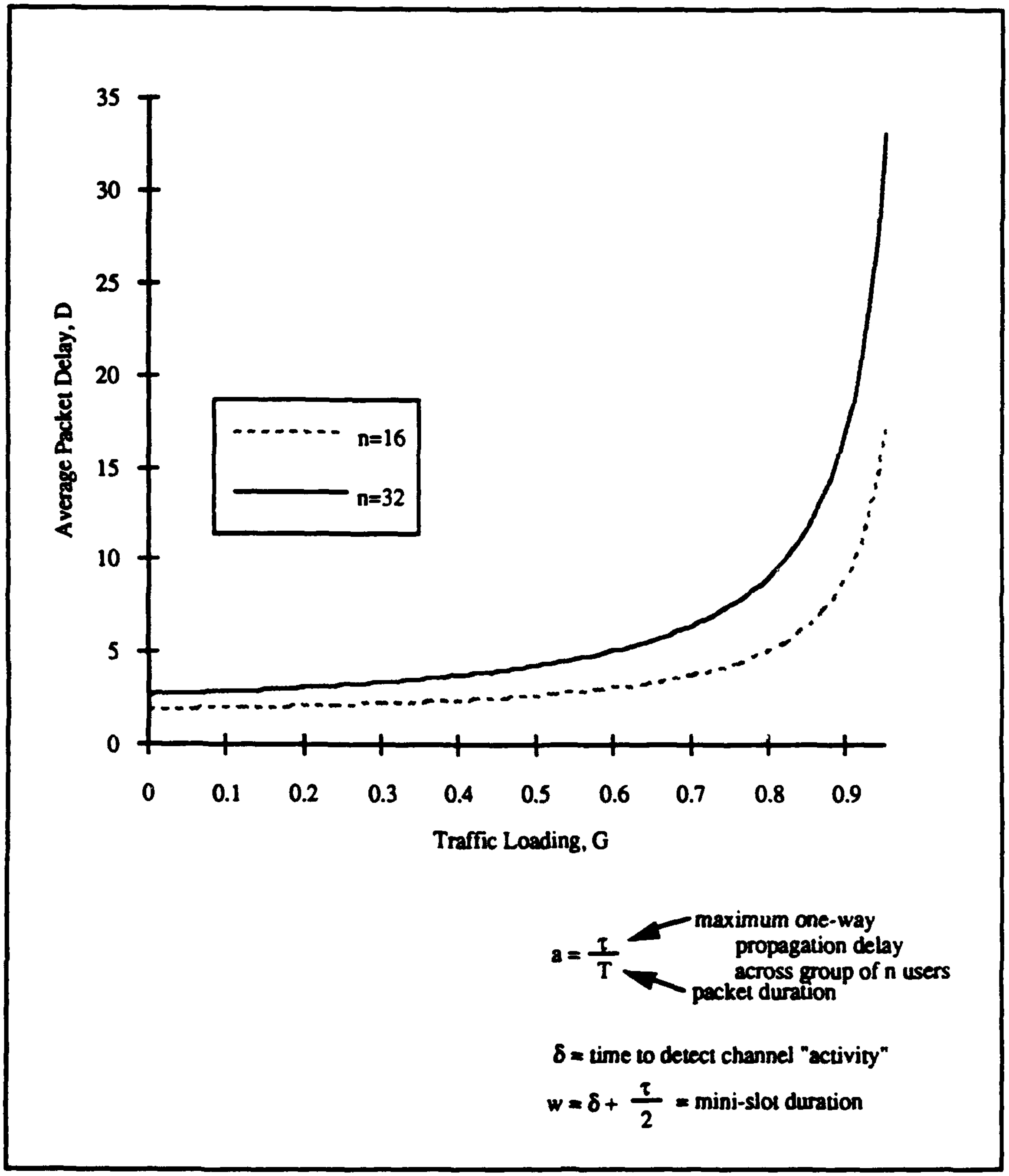


Figure 11-8. Average Packet Delay vs. Traffic Loading

Then the channel utilization, U, can be calculated from equations (16), (21b) and (23), e.g., see Table 11-1.

Table 11-1. Channel Utilization, U, for Several Values of Traffic Loading, G ($a = 0.1$; $\delta/T = 0.05$, i.e., $v = 0.1$)

G	Q/n	z	s	U(%)
0.3	0.143	0.0018	1.41	1.26
0.6	0.25	0.0209	2.29	8.36
0.9	1	0.534	4.66	53.4

Note that despite the fact that S = G, there is considerable overhead on the channel due to (idle) mini-slots, even at relatively high values of traffic loading. High utilization tends to imply very high delay.

This protocol is stable since there always will be finite throughput no matter how much backlog accumulates in the queues of users not transmitting.

11.3.5 Non-Persistent and p-Persistent CSMA

Recall that in Carrier Sense Multiple Access (CSMA) schemes, a user u with a packet to transmit first senses whether any other user is transmitting. If not, various strategies can be employed by user u [9]. Different packets arriving at user u are treated in order of their arrival in the following fashion. Of course, if user u has no packet to transmit, no action is taken by that user.

In non-persistent CSMA, user u transmits if the channel is sensed idle and it has a packet to transmit; if the channel is sensed busy, user u reschedules the packet transmission according to some predetermined retransmission delay distribution, and at that time the process is repeated for that packet.

In 1-persistent CSMA, user u transmits with probability 1 if the channel is sensed idle and it has a packet to transmit; if the channel is sensed busy, user u waits until the channel becomes idle and then immediately transmits the packet with probability 1.

The non-persistent and 1-persistent CSMA protocols can be applied in either a non-slotted or slotted form. Of course, if channel time is slotted, all users must have accurate time. In this case, channel time is partitioned into mini-slots of duration τ, the maximum one-way propagation delay among the users, and each user must know time accuracy to a fraction of a mini-slot. When a packet is ready for transmission, it is actually transmitted at the beginning of the next mini-slot.

In p-persistent CSMA, a mini-slotted scheme, user u transmits at the beginning of the next mini-slot with probability p if the channel is sensed idle and it has a packet to transmit; the packet is delayed one mini-slot with probability 1-p. If the channel is still idle, the same process is repeated; otherwise, user u reschedules this packet's transmission according to some retransmission delay distribution, and at that time the process is repeated for that packet. If the channel is sensed busy originally, user u waits until the channel becomes idle and repeats the whole process.

The throughput of these CSMA protocols was shown in figure 11-2 for $a = \tau/T = 0.01$, where T is the packet length, and several values of p. Referring to that figure, note that non-persistent CSMA and p-persistent CSMA with $p \leq 0.1$ can provide more than twice the throughput of Slotted ALOHA with an a as small as 0.01. As a increases to 0.1, this advantage of CSMA gradually decreases to about 50% more throughput [9, fig. 5, p. 474, 10].

A comparison of the (normalized) throughput and delay of the ALOHA and CSMA schemes is shown in figure 11-9 for a = 0.01 and the optimum value of p based on published simulation results [9, fig. 4, p. 473; 10]. Again, the delay of the CSMA protocols would increase significantly as a approaches 0.1.

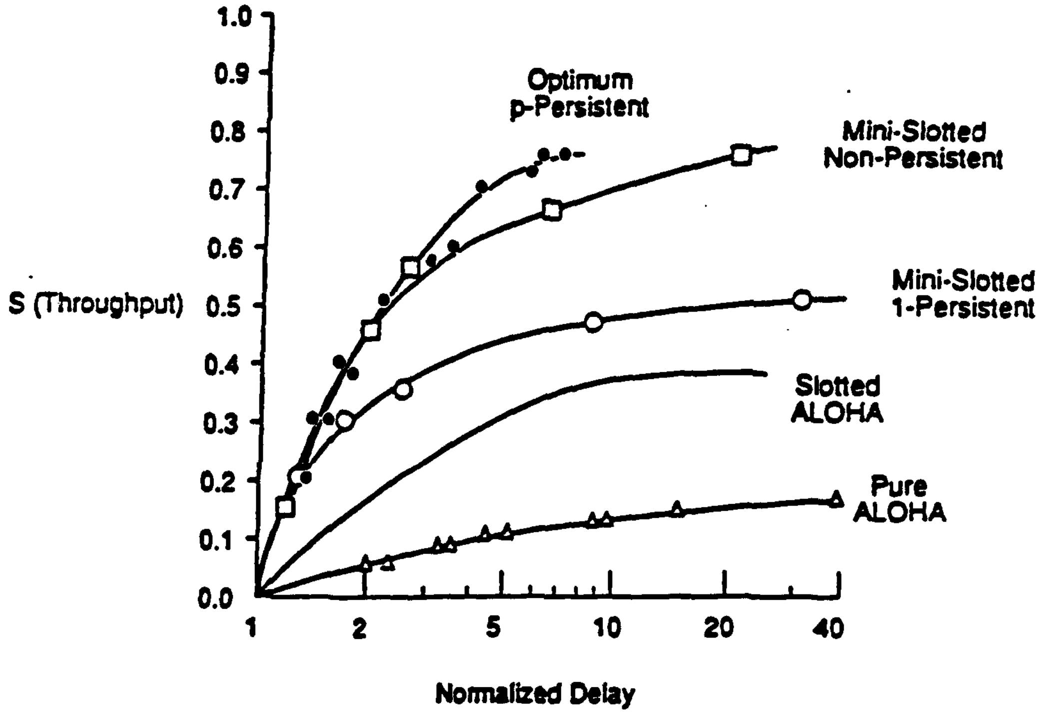


Figure 11-9. CSMA and ALOHA: Throughput-Delay Tradeoffs from Simulation
(Normalized Propagation Delay a = 0.01)

The value of p in p-persistent CSMA can be varied in response to observations of current channel loading conditions. For example, near channel "saturation" when the throughput is near the maximum, but dropping, p can be decreased (increased) as packet collisions on the channel become more (less) frequent. When such a capability is provided, this is called adaptive p-persistent CSMA. This adds some stability to the basic scheme which is unstable, cf., figure 11-2.

In the VDR development, the currently selected RA scheme is a non-adaptive p-persistent CSMA protocol with $p = 0.5$ for ground stations and $p = 0.05$ for aircraft. The mini-slot duration is approximately 2 ms for the Offset Quadrature Amplitude Modulation (OQAM) selected. The specific value is tied to the duration of the "pre-key" sequence, not the maximum one-way propagation delay which is unspecified in the AVPAC system.

11.3.6 Virtual Time CSMA

In addition to a real-time clock, this user-autonomous algorithm also employs a "virtual-time" clock [15] which, in general, lags behind the real-time clock. As will be shown later, the reason for this is to attain improved throughput and delay performance under certain conditions compared to other RA schemes such as Slotted ALOHA and 1-persistent CSMA.

Initially the two clocks show the same time. As messages arrive, their arrival times on the real-time clock are noted. Each message is transmitted as soon as the virtual-time clock reads the same as that message's arrival time. The (broadcast) channel is busy when at least one message is being transmitted; if two or more messages overlap in time, of course, a collision occurs. The virtual-time clock is stopped when the channel is busy. During idle (non-busy) periods the virtual clock is speeded up by a factor $\eta > 1$ until another message is transmitted or it catches up with the real-time clock, after which the virtual-clock runs with the real-time clock until the next time a message is transmitted. It is this speed-up process that allows more efficient use to be made of the channel when the maximum propagation delay, τ between the users is much less than a message duration, taken here as a fixed time, T, corresponding to our single packet duration. As usual the parameter $0 < a = \tau/T < 1$ is used to represent this relative propagation delay.

The performance of the virtual-time-algorithm can also be improved if the collision detection feature is feasible. In this case, if any transmitting user can sense a collision with another user's transmission, it does so, and then the first user immediately "jams" the channel with a sufficiently high-power but relatively short duration signal that all other users will likely then recognize the collision, as well. Let the parameters $0 < c < 1$ and $0 < c' < 1$ represent the relative lengths of the transmission time for a colliding message and the collision plus jamming time, respectively.

Thus, for synchronous (slotted) operation on which we will concentrate here for brevity, idle slots last for relative time a. A successful-transmission busy-slot lasts for $a + 1$, and a collision busy-slot lasts for $a + c$. We will assume collision detection but will ignore the potential collision-plus-jamming parameter c', assuming any jamming is part of the parameter c. Asynchronous operation is also possible with Virtual Time CSMA, although performance is not as good as with synchronous operation. As with other RA algorithms, synchronous operation requires sufficiently accurate local time and an increase in operational complexity. In this case users must be synchronized to a fraction of an idle slot.

Throughput

In any event, for synchronous operation the equilibrium, i.e., "steady state", throughput S for Virtual Time CSMA can be expressed as the average of the throughputs, S_1 and S_2, for each of two possible modes, to be defined shortly, divided by the average of the expected slot durations, $E(L_1)$ and $E(L_2)$, for the same two modes:

$$S = \frac{\sum_{m=1}^{2} S_m p_m}{\sum_{m=1}^{2} E(L_m) p_m} \tag{24}$$

Mode 1 ($m = 1$) is active when the virtual-time clock is aligned with the real-time clock or when the channel is busy with the virtual-time clock stopped but the virtual-time clock had just been aligned. This mode occurs with equilibrium probability p_1. Mode 2 ($m = 2$) is active when the virtual-time clock is speeded up by the factor η or when the channel is busy

with the virtual-time clock stopped but the virtual-time clock had just been behind the real-time clock. This mode occurs with equilibrium probability p_2. Note that the throughput of equation (24) is generalized from the usual concept of throughput in Slotted ALOHA, for example, where all the slots are identical in duration. Here the slots are of three different durations all under two distinct probability distributions depending on which of the two modes of operation is active.

Before expanding upon equation (24), let us briefly review the standard theoretical model underlying all the RA algorithms discussed in this memorandum. An infinite number of users and a Poisson message arrival distribution are assumed.

An infinite number of users is pessimistic in that the theoretical results for expected throughput and delay would improve for a finite number of users. We assume an infinite number of users, and an aggregate new and old (messages that need retransmission because of previous collisions) packet arrival rate of $\lambda = G/T$ packets per unit time, with Poisson probability distribution. Recall that for the Poisson distribution of equation (0), $E(k) = \sigma_k^2 = G$ for an interval of duration $t = T$. Now specializing to equation (24), we have

$$S_1 = aGe^{-aG} \tag{25a}$$

$$S_2 = a\eta Ge^{-a\eta G} \tag{25b}$$

the probability that exactly one packet arrives in time a and in time $a\eta$ in Modes 1 and 2, respectively. Similarly,

$$\begin{aligned} E(L_1) = ae^{-aG} &+ (a+1)aGe^{-aG} \\ &+ (a+c)\left[1-(1+aG)e^{-aG}\right] \end{aligned} \tag{26a}$$

$$\begin{aligned} E(L_2) = ae^{-a\eta G} &+ (a+1)a\eta Ge^{-a\eta G} \\ &+ (a+c)\left[1-(1+a\eta G)e^{-a\eta G}\right] \end{aligned} \tag{26b}$$

The three terms in equations (26a) and (26b) correspond to zero packets, exactly one packet, and more than one packet, in the three slot durations, a, a+1, and a+c, respectively.

The equilibrium probabilities p_1 and $p_2 = 1 - p_1$ are obtained as follows by equating the denominator of equation (24) with the average advance of the virtual-time clock, namely, $ap_1 + a\eta p_2 = a(1-\eta)p_1 + a\eta$. Using equations (26a) and (26b), the result is

$$p_1 = \begin{cases} 0, & \text{if } x \leq 0 \\ x, & \text{if } 0 < x < 1 \\ 1, & \text{if } x \geq 1 \end{cases} \tag{27a}$$

where

$$x = \frac{E(L_2) - a\eta}{E(L_2) - a\eta - [E(L_1) - a]} \tag{27b}$$

Figure 11-10 shows the throughput, S, versus total message traffic loading, G, for several values of the parameters a, c, and η. It has already been seen in the background section that interesting values of a range from about a = 0.01 to about a = 0.4. We shall assume that a collision can be detected in no more than 100 bits. Again, from the background section, packet lengths range from an assumed minimum of L_{min} = 200 b to an assumed maximum of about L_{max} = 10,000 b. Thus, c ranges from about c = 0.01 to c = 0.5.

Delay

The normalized expected equilibrium message delay is given by

$$D = D_o + (G/S - 1)\, D_r + (a + 1) \tag{28}$$

where D_o is the average initial delay from the arrival of a new message until its first transmission attempt (or in the case of non-persistent CSMA, until the decision is made not to transmit the message), D_r is the average retransmission delay from the start of one unsuccessful attempt to transmit a message until the start of the next attempt, and the term a + 1 represents the propagation delay and transmission time for the final (successful) attempt.

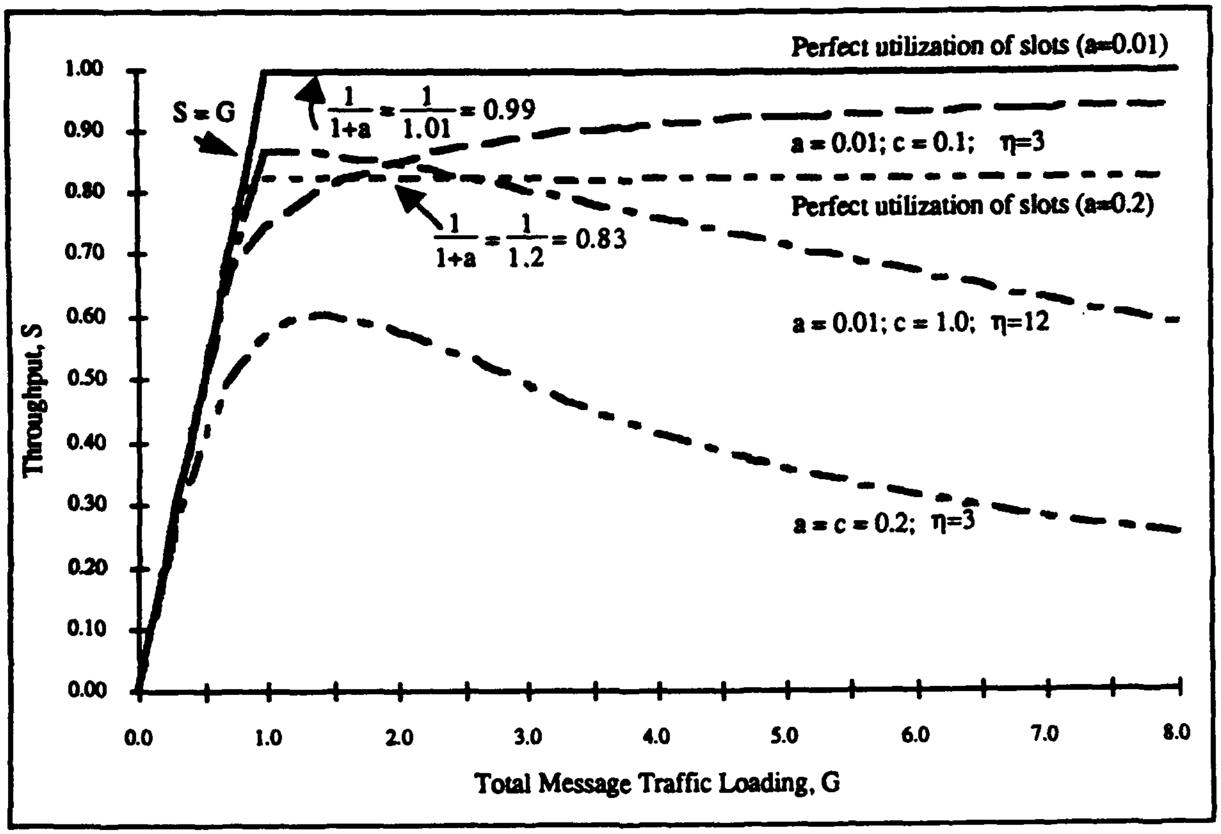


Figure 11-10. Throughput vs. Traffic Loading for Virtual Time CMSA

The factor (G/S -1) represents the expected number of retransmission attempts per successful message, i.e., in some arbitrarily long time, T_a, there are an expected number of ST_a successful message transmissions out of a total of GT_a expected message (new and old) transmissions. Thus, the relative number of unsuccessful transmissions is

$$\frac{GT_a - ST_a}{ST_a} = \frac{G}{S} - 1 \tag{29}$$

The initial delay, D_0, can be estimated by assuming a reasonable queueing model [16], the details of which are beyond the scope of this paper, with the following parameters:

ω = $a\eta$, normalized elementary time unit of discrete time M/G/1 queue [17, p. 31]; no more than one packet can arrive (or depart) at any discrete time $k\omega$, $k = 0, \pm 1, \pm 2, \ldots$.

W = average normalized waiting time in M/G/1 queue

$\frac{1}{\mu}$ = average of normalized service time distribution

σ^2 = variance of normalized service time distribution

λ = G, average normalized (Poisson) packet arrival rate

ρ = $\frac{\lambda}{\mu} = \frac{G}{\mu} < 1$, traffic intensity

Molle [18] has shown that

$$W = \frac{1}{\mu} \frac{\rho\left(\mu^2\sigma^2+1\right) - \left(1 - e^{-a\eta G}\right)}{2(1-\rho)} \tag{30}$$

The normalized initial delay, D_0, for this model can be expressed as

$$D_0 = \frac{\omega}{2} + \frac{W(\eta-1)}{\eta} \tag{31}$$

(refer to [17] for details) where η is the speed-up factor of Virtual Time CSMA.

The retransmission time for the synchronous case is

$$D_r = a + c + r \tag{32}$$

where $r >> 1/(\eta-1)$ is the average time it takes for the virtual clock to advance by the scheduled retransmission delay. For purposes of computation we shall assume that

$$r = 10/(\eta-1). \tag{33}$$

Combining equations (28), (32), and (33), the total delay of equation (28) becomes

$$D = \frac{a\eta}{2} + W\frac{(\eta-1)}{\eta} + \left(\frac{G}{S}-1\right)\left(a + c + \frac{10}{\eta-1}\right) + a + 1 \tag{34}$$

Figure 11-11 shows this synchronous Virtual Time CSMA delay for two sets of parameters.

Virtual time CSMA can be made stable by adjusting the virtual-time clock rate, η, dynamically. (See p. 923 of Molle [15] between equations (5) and (6).)

11.4 TRADEOFFS

In comparing RA schemes the principal question is the extent to which they support real-time operations. Put another way, what is the range of throughput, or more specifically, what is the maximum throughput that can be sustained without exceeding a given delay? For the purposes of comparison we have set 200 ms, our nominal definition of real-time, 1s, near-real-time, and 5s, non-real-time, as delay thresholds to meet.

The first step is to compute the range of maximum propagation delay, τ, for each of the three altitude regimes using the propagation delay to the subaircraft point and to the radio horizon as the lower bound and upper bound, respectively (see Table 11-2).

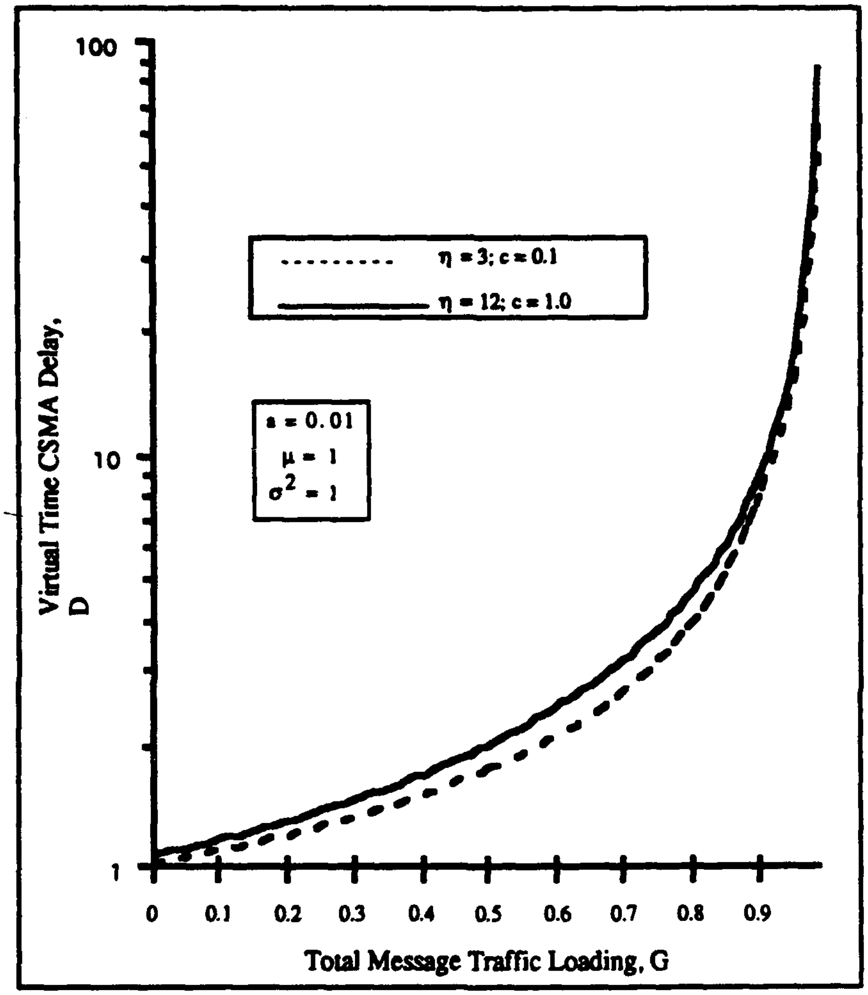


Figure 11-11. Delay for Synchronous Virtual Time CSMA

Table 11-2. Extremes of Maximum Propagation Delay, τ, for Three Altitude Regimes (h is altitude, in feet; c is speed of light in feet per second)

h	$\tau_{min} = h / c$	$\tau_{max} = 5280 \sqrt{2h} / c$
72,000 ft	73.1 μs	2040 μs
18,000 ft	18.3 μs	1020 μs
4,500 ft	4.6 μs	509 μs

The next step is to select interesting combinations of R and L to compute a representative set of packet durations, T = L/R, and the corresponding expected packet delays, in units of packet slots, that satisfies the real-time criterion of 200 ms. The latter can be thought of as the threshold for real-time performance (see Table 11-3).

Table 11-3. Packet Durations, T = L/R, of Interest and the Real-Time Thresholds

R	L	T = L/R	Number of Packet Slots in 200 ms
20 kb/s	256 b	12.8 ms	15.6
	512 b	25.6 ms	7.8
	1024 b	51.2 ms	3.9
	2048 b	102 ms	2.0
	4096 b	205 ms	1.0
	8192 b	410 ms	0.5
40 kb/s	256 b	6.4 ms	31.3
	512 b	12.8 ms	15.6
	1024 b	25.6 ms	7.8
	2048 b	51.2 ms	3.9
	4096 b	102 ms	2.0
	8192 b	205 ms	1.0

Now each of the RA schemes will be assessed, in turn, as to their capability to support real-time communications. Keep in mind that even though the <u>expected</u> packet delay may satisfy the real-time threshold, a large proportion (the exact fraction dependent upon the

probabilistic delay distribution) of the packets would be (successfully) delivered after the real-time threshold of 200 ms. This implies that to provide higher assurance that most packets would be delivered in real-time, the packet duration T = L/R must be decreased significantly.

11.4.1 Pure ALOHA

Referring to figure 11-2, it appears that a reasonable level of throughput, say S = 0.125, can be sustained at a channel loading of about G = 0.18, cf., equation (1a). From figure 11-3a we see that this channel loading implies an expected delay of about D = 4 or 5. Examining equation (2) for the values displayed in figure 11-3a, we see that, for example, $D = (3+4.5)\left(\frac{0.18}{0.125} - 1\right) + 0.1 + 1 \approx 4.4$ for the case of P = 1, K = 10, corresponding to a medium delay curve. The particular value of $a = \tau/T$ is negligible for any of the curves. Looking at Table 11-3 we conclude that for real-time operation, the packet lengths should not exceed L = 1024 b at R = 20 kb/s and L = 2048 b at R = 40 kb/s; to be safer packets half that long should be used.

11.4.2 Slotted ALOHA

In a similar way, an S = 0.25 and a G = 0.36 imply about the same delay, and we draw the same implication about the maximum packet lengths for real-time operation. The advantage of Slotted ALOHA over Pure ALOHA is higher throughput for the same delay, or less delay at the same throughput.

11.4.3 Reservation ALOHA

Referring to figure 11-4, if we scale the delay by a factor 2.5 = 50 kb/s ÷ 20 kb/s, then the minimum delay (at <u>zero</u> throughput) for FIFO with Slotted ALOHA reservations is 0.6 s × 2.5 =1.5 s. Thus, not even near-real-time operation can be supported; only non-real-time, where the delay threshold is 5 s.

11.4.4 Collision Resolution Algorithms

As already pointed out in the Issues section, D = 8 corresponds to a throughput in the vicinity of S = 0.25 for the basic CRA algorithm. Again, the parameter a is not an influential factor. From Table 11-3 this suggests that L should not exceed 512 b for R = 20 kb/s and 1024 b for 40 kb/s, if real-time performance is to be expected. This is a factor of two disadvantage compared to Slotted ALOHA.

11.4.5 Mini-Slotted Alternating Priorities

From Table 11-1 we see that the channel loading, or in this case the throughput, must be in the vicinity of G = S = 0.9 for a reasonably efficient utilization of the channel. From figure 11-8 this implies an expected delay of about D = 8 or 9 for n = 16 users and D = 18 or 20 for n = 32 users. Once again, returning to Table 11-3, real-time operation for n = 16 is limited to about L = 512 b for R = 20 kb/s (and 1024 b for 40 kb/s). Only near-real-time and non-real-time operations are possible at 20 kb/s for n = 32; at 40 kb/s only about L = 512 b is possible for real-time operation.

In contrast to the previous RA schemes, the value of the parameter a can have a significant influence in the delay at the higher values of throughput. The above results were obtained for a relatively large a = 0.1 and a "balanced" value of $\delta T = 0.05$*, cf., equations (21b) and (22). For n = 16, L = 512 b and R = 20 kb/s, the maximum tolerable real-time packet duration of 25.6 ms, cf., Table 11-3, and the altitude regimes of Table 11-2, yield the range of values $0.02 \leq a \leq 0.08$. Note that a = 0.1, the value assumed in figure 11-8, is near the upper end of this range.

11.4.6 p-Persistent CSMA

For non-adaptive p-persistent CSMA with a minislot duration of $\tau = 2$ ms, for example, and an a = 0.01, so that figure 11-9 can be used, we have T = 200 ms. However,

* Note that this corresponds to detecting the absence of a signal on the channel in only 25 b for n = 16 and 20 kb/s, and for n = 32 and 40 kb/s.

from Table 11-3, for R = 20 kb/s and this packet length of L = 4096 b we see that the expected packet delay must be about D = 0.5 for real-time operation in this case. But figure 11-9 shows that no throughput can be sustained; thus, only near-real-time and non-real-time operation is possible for this set of values.

Now suppose a = 0.04 instead. Then T = 50 ms and the expected delay must be no larger than about D = 2. In this case figure 11-9 suggests that as much as 45% throughput is possible for optimum p-persistent CSMA. However, that curve would be lower for a = 0.04 compared to a = 0.01, as indicated in the issues section. A rough estimate would be S = 0.3 or 0.35 for D = 2 and a = 0.04. Simulation is recommended for a more accurate estimate of the maximum throughput for real-time operation; anyway, the maximum packet length is limited to 2048 b for R = 20 kb/s. We also note that the minislot of 2 ms corresponds with the maximum propagation delay from the highest altitude regime, cf., Table 11-2; it is not clear just how the 2 ms minislot duration was selected in the design of the VDR.

For a = 0.04 and R = 40 kb/s, we have the maximum tolerable L = 4096 b, and the need for D = 2, also, for real-time operation.

11.4.7 Virtual Time CSMA

For virtual time CSMA with n = 12 and no collision detection (c = 1.0), the "knee" of the delay curve of figure 11-11 suggests a channel loading of about G = 0.8 and a D ≈ 5. The corresponding throughput on the n = 12 curve of figure 11-10 is just below G, say S ≈ 0.75. In this case the results are rather insensitive to the parameter a. It appears that $a = \tau/T$ can be as large as a = 0.1 without much effect on delay but with some drop in throughput, perhaps to about 60 to 70%, an estimate deduced simply by a gross attempt to interpolate the curves of figure 11-10 for G ≈ 0.8. Again, additional detailed calculations would be required for a more accurate estimate.

Referring to Table 11-3, a D of 4 or 5 implies that for real-time operation, the maximum packet length is L = 512 b for R = 20 kb/s and 1024 b for 40 kb/s. Because this implies a packet duration of 25.6 ms, and because the highest altitude regime is associated with a 2 ms propagation delay, cf., Table 11-2, $a \approx \tau/T = 0.08$ is assumed. Therefore, users must synchronize to an accuracy of about $0.1 \times 0.08 \times 25.6$ ms = 0.2 ms.

11.4.8 Summary

Table 11-4 summarizes the foregoing discussions as to whether each RA scheme considered can support real-time operations, and, if so, the maximum throughput that can be sustained, as well as any constraints that apply, particularly, the maximum packet length recommended. The impact of the parameter $a = \tau/T$, the stability of the basic algorithm depicted, the required accuracy, and an indication of the relative complexity of the schemes are also noted.

11.5 IMPACT/IMPORTANCE

11.5.1 Best RA Scheme Performers

If the primary objective is to support real-time operations with high throughput, the RA scheme called Virtual Time CSMA appears to be the clear winner. The tradeoff between throughput and expected packet delay outlined in the previous section yields the profiles of figure 11-12. Virtual Time CSMA can support throughputs of up to approximately 70% with packet durations no longer than 12.8 ms (L = 256 b at R = 20 kb/s and L = 512 b at R = 40 kb/s). As the maximum sustainable throughput decreases, longer packets can be employed as shown by the "staircase" profiles in figure 11-12; at the lowest indicated throughput of approximately 15%, average packet durations of 205 ms are possible.

Optimum p-persistent CSMA yields the next best throughput-delay profiles, after Virtual Time CSMA. Referring to figure 11-12, Virtual Time CSMA sustains approximately 50% more throughput at an average packet duration of 102 ms, and roughly 25% more throughput at average packet durations of 51.2 ms, 25.6 ms and 12.8 ms.

Real-time operation cannot be supported at all with packet durations longer than 205 ms with Virtual Time CSMA, and 102 ms with optimum p-persistent CSMA. In addition, one should keep in mind that all these results are based on expected packet delay, i.e., a significant portion of the packets of a given fixed length in figure 11-12 will take longer than the real-time threshold of 200 ms to arrive correctly. A much more detailed analysis

Table 11-4. Characteristics of the RA Algorithms Discussed With Emphasis on Their Ability to Support Real-Time Operations

RA Scheme	Maximum Packet Length For Real-Time Operation 20 kb/s	40 kb/s	Sustainable Throughput	Impact of Parameter a *	Stability	Required Timing Accuracy	Complexity
Pure ALOHA	512 b	1024 b	0.125	negligible	unstable	none	lowest
Slotted ALOHA	512 b	1024 b	0.25	negligible	unstable	~1ms	low
Reservation ALOHA	- - - - - - - - - - - Cannot support real-time or near-real-time operations, only non-real-time - - - - - - - - - - -						
Collision Resolution Algorithms	256 b	512 b	0.25	negligible	stable	~1ms	medium
Mini-Slotted Alternating Priorities	512 b (n=16) - - - (n=32)	1024 b 512 b	0.53 (utilization)	significant (a=0.1)	stable	0.13 ms	medium high
p-Persistent CSMA	2048 b	4096 b	0.3 to 0.35	significant (a=0.04)	unstable	0.2 ms	medium
Virtual Time CSMA	512 b	1024 b	0.6 to 0.7	insensitive (a=0.08)	unstable	0.2 ms	highest

* $a = \tau/T$, ratio of maximum one-way propagation delay to packet duration

n = number of users

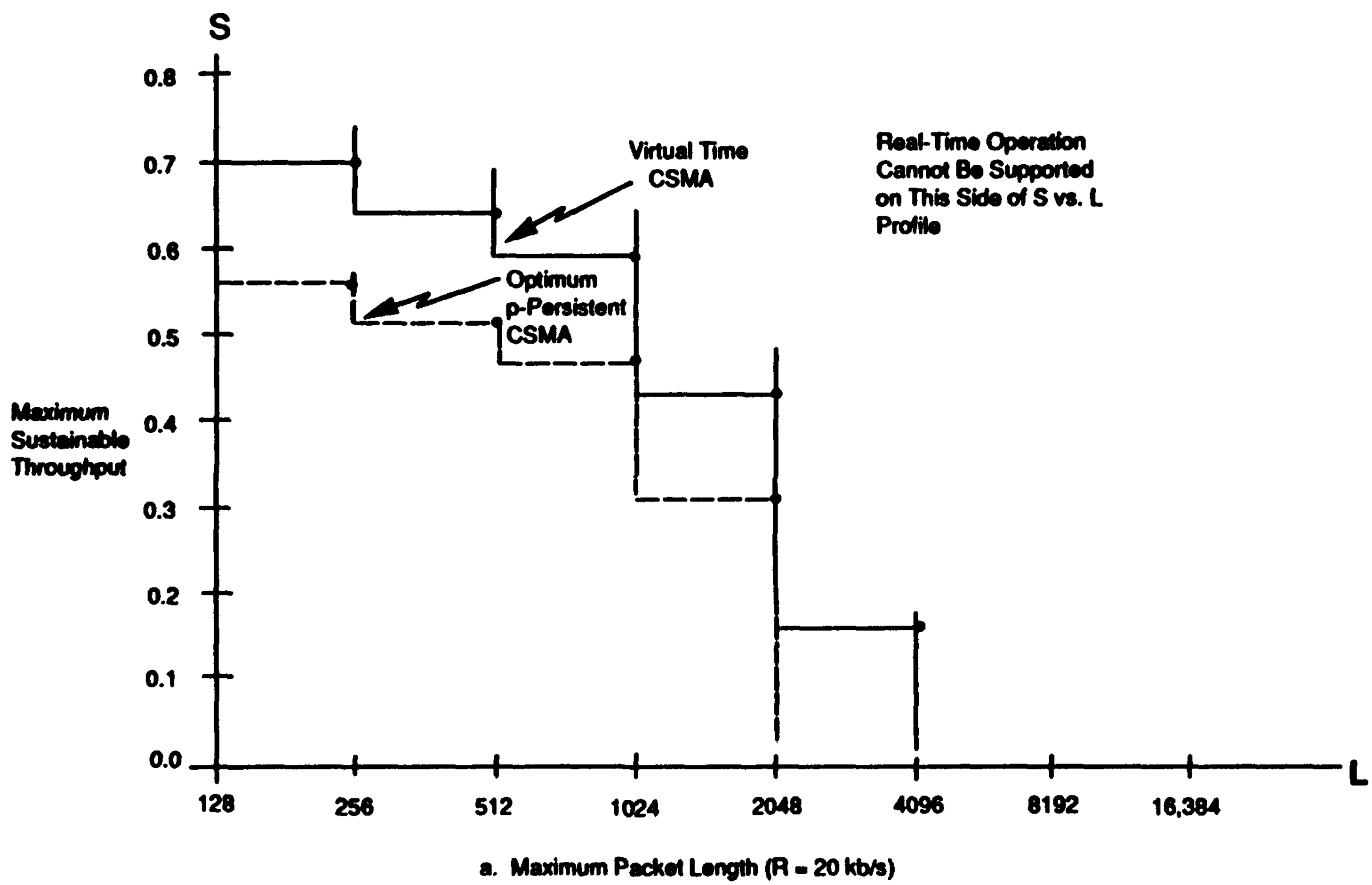


Figure 11-12a. Approximate Regions of Support to Real-Time Operations by the Most Capable Random Access Schemes

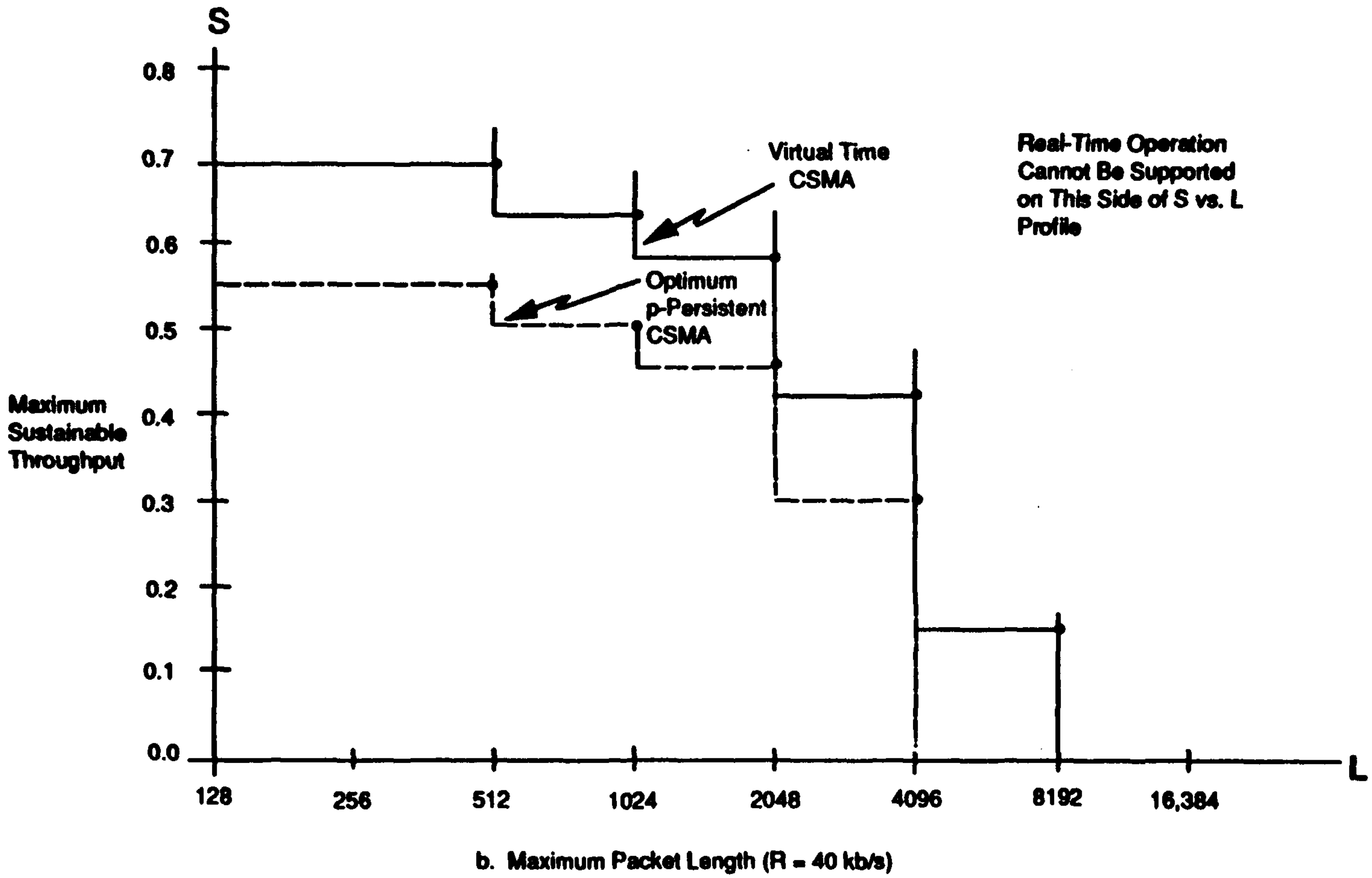


Figure 11-12b. Approximate Regions of Support to Real-Time Operations by the Most Capable Random Access Schemes

would be required to generate a probabilistic distribution of packet delay as a function of packet duration, for a given throughput level, and for each RA scheme. It is observed that much shorter packet durations than indicated by these averages are necessary if a high percentage of packets are to be received correctly within the real-time threshold.

Related Analysis of p-Persistent CSMA

A study by D'Amours and Mazur [19] dealing with this issue for p-persistent CSMA and the VDR recently came to our attention. For various reasons which they explain in describing their model, the authors state that their results are optimistic: the "maximum" throughputs are upper bounds and the "maximum" delays are lower bounds. They also suggest that detailed simulations should be performed to get more accurate results about the situations of principal interest. Nevertheless, their paper provides useful information that suggests how much greater the maximum delay might be compared to the average delay. They give numerical examples where the maximum delay is only exceeded 5% of the time, i.e., they compute the delay which is not exceeded with 95% confidence, and compare it with the average (50% confidence delay).

For important nearly optimal (good value of p) cases where the system is operating at high throughput values, they estimate that the 95% delay is more than six times the 50% delay. This suggests that the packet lengths of Table 11-4 and figure 11-12 would have to be several times shorter to be able to support real-time communications if the 200 ms threshold were not to be exceeded 95% of the time. Unfortunately, it is dangerous to estimate the quantitative impact of this 95% confidence measure of delay, even on just our p-persistent CSMA model, without performing considerable additional work. We feel this is beyond the scope of our present effort which was limited to estimating the average (50%) delay.

It is noted that the average packet length of 30 bytes (240 b) used by D'Amours and Mazur in their study is near the lower end of the range of 200 b to 10,000 b assumed in our work. According to their references this 240 b average was used in considering VHF data link access protocols for ATC applications under the auspices of Working Group C of the AMCP. However, we have been told [20] that the VDR will accommodate packets as long as 1024 bytes (8192 b). It is clear from Table 11-4 and figure 11-12 that none of the RA

schemes discussed thus far can support real-time operations, even with the average delay (50% confidence level) definition, at significant throughput levels if packet lengths are greater than 2048 b at 20 kb/s and 4096 b at 40 kb/s, i.e., if packet durations are greater than 102 ms.

D'Amours and Mazur noted that the average packet duration would be 240 b/21 kb/s = 11.4 ms with the VDR's 4-OQAM modulation, and 240 b/42 kb/s = 5.07 ms for 16-OQAM. They used an overall average packet duration of (11.4 + 5.07) ms/2 = 8.24 ms, assuming that each of the 4-OQAM and 16-OQAM modes would be used half the time. It is our understanding [20] that the 16-OQAM mode would predominate, and that 4-OQAM would only be used for synchronization and channel equalization, where every 16th symbol of the otherwise 16-OQAM data stream would be a 4-OQAM symbol to help measure amplitude fluctuations on the channel that may be caused by fading. Thus, most of the packets would be only about 5 ms in duration. Since propagation delays can easily be at least 1 ms, cf., Table 11-2, this means that the p-persistent CSMA scheme with an $a \geq 0.2$ may be somewhat inferior than even the optimistic bounds portrayed by D'Amours and Mazur using their $a = 0.125$.

As these authors further noted in their study, 16-OQAM modulation would not perform very well in a multipath fading environment. This could lead to additional packet retransmissions beyond those attributed to packet collisions. In this respect their analysis is even more optimistic with respect to the VDR.

As we have assumed throughout, and as D'Amours and Mazur assume in their study, acknowledgment packets are returned to the transmitter over a separate channel without error. If this were not the case, as D'Amours and Mazur point out, less throughput and greater delay would be experienced on the forward channel because additional transmissions of old packets would be required. In the VDR development the uplinks and downlinks share the same frequency channel [20]. This means that acknowledgment packets must compete for the same broadcast channel resource as the forward packets. This leads to a further degradation in throughput and increased delay as could be expected on the forward channel if a separate frequency channel were used for the uplinks.

It has been shown that greater overall spectral utilization can be achieved if the uplink and downlink frequencies are separated [21, 22]. Although two frequencies are required for each channel instead of one, more than a factor of two gain in frequency reuse is possible, even in the present operational environment of the current frequency plan. Theoretically, a gain of as much as four in frequency reuse is possible with a uniform distribution of aircraft and ground radio sites.

D'Amours and Mazur present a plausible method for computing the statistical delay for an arbitrary confidence level using p-persistent CSMA. The method depends on the statistical independence of two random variables, the number of transmissions required to successfully transmit a packet, and the average number of mini-slot deferrals per transmission. Although not obvious, this independence seems reasonable because a decision to defer to the next mini-slot only depends on whether the channel is busy; however, some of the time, a busy channel would be the result of packet collisions of two or more users attempting to transmit in the next mini-slot after sensing the channel idle in the previous mini-slot. Such collisions obviously increase the number of transmissions required. Therefore, it seems that there would be at least a weak statistical dependence between the two random variables. The authors also assume that the two random variables affect the delay equally. It is not clear that this necessarily would be the case. Thus, although the results presented in their study are plausible, they should be viewed as optimistic, as the authors state, and simulations should be performed to validate the model employed.

An even more recently received addition to the analysis of D'Amours and Mazur [23] shows that the p-persistent CSMA delays to be expected in the VDR will be even greater than discussed above. Their original analysis had ignored the "training sequence". Consequently, the packet durations are longer, and the lower bounds on the expected 50% and 95% delays are actually higher than given earlier. We will not attempt to quantify the differences since we do not agree entirely with their model, as outlined above.

In the attachment [23] to their original analysis, D'Amours and Mazur state:

- AOC packets being considered for AVPAC are much longer than those for ATC. [ATS, presumably; they assumed 240 b for the average ATC message in their original analysis.] Thus, there would be more packets contending for the channel following transmissions of AOC packets as well as the increased waiting times. This would result in more packet collisions and longer "average" [50%] and "maximum" [95%] delays. This requires even more reduction in throughput to satisfy maximum delay conditions [real-time requirements].

- "... it becomes unclear whether or not AVPAC provides any capacity improvements over current technology."

Other Characteristics of Best Performing RA Schemes

Although Virtual Time CSMA outperforms p-persistent CSMA, it is a somewhat more complex RA scheme. It could be that the superior performance envelope of Virtual Time CSMA is not worth the added complexity. Evidently this was the case with the VDR development because Virtual Time CSMA was considered initially [20], but a p-persistent CSMA scheme was selected. Both RA schemes require an absolute timing accuracy of about 0.2 ms, and both are unstable. Although Virtual Time CSMA appears to be more insensitive to the parameter a, this is not a determining factor in deciding between the two schemes for a VHF A/G communications environment.

11.5.2 Assessment of Lower Performance RA Schemes

Recall that mini-slotted alternating priorities requires a relatively large channel loading to achieve a reasonably high utilization of the channel. Thus, one would expect a considerable variation in performance with this scheme, and although it is stable, mini-slotted alternating priorities is not a very attractive method for sustaining high throughput with relatively small delay.

If lower values of sustainable throughput in the order of 25% are acceptable, then Slotted ALOHA becomes more attractive because of its low complexity and more relaxed absolute timing accuracy requirement of only about 1 ms. Again, however, real-time operation only can be supported for packet durations of no more than about 25.6 ms at

Slotted ALOHA's best throughput values. Referring to Table 11-4, CRAs yield approximately the same throughput as Slotted ALOHA but can only support real-time operation with half the average packet lengths. Although CRAs are stable, Slotted ALOHA can be made more stable with the slight modification described in the Issues section. Finally, despite the fact that Pure ALOHA requires no absolute timing, and is therefore the simplest of all the RA schemes, roughly only half the throughput of Slotted ALOHA is possible. Because the level of complexity of Slotted ALOHA seems quite acceptable, Pure ALOHA would most likely be rejected in favor of another scheme in most applications today.

Reservation ALOHA has little chance of supporting real-time operation because it is employed in tandem with another access scheme like TDMA. However, in the context of hybrid access techniques, it would be worthwhile to explore the use of Reservation ALOHA in demand assigned access schemes. See branchpoint 4.2.2 in the appendix. This may be conducted in future work.

11.6 TRANSITION

Although this section is included to be consistent with the format of other decision tree papers, it is probably the least important section of this paper because RA schemes would generate relatively few, if any, problems when implementing evolutionary changes in VHF A/G communications systems. Nevertheless, we have included the following remarks for consideration by decision makers.

One of the most important issues for any multiple access protocol scheme is whether it is compatible with the Open Systems Interconnection (OSI) model of the International Standards Organization (ISO). In the field of air ground (A/G) communications this issue is akin to evaluating the extent to which the access protocol is compatible with the Aeronautical Telecommunications Network (ATN) [1]. This question is particularly appropriate for packet-oriented communications systems like the RA schemes discussed in this memo. The point is that the more an RA scheme is or can be aligned with the ATN concept, the easier may be the transition period of introducing the RA scheme into operational A/G communications systems.

As we have seen the RA algorithms discussed cover a fairly broad range of complexity (see Table 11-4 of the tradeoffs section) that is most easily thought of in terms of software as opposed to hardware. All the RA schemes discussed are well established in the literature, if not in the field. There are no apparent difficulties that would prevent any of them from being used within the ATN. In fact, the VHF Data Radio (VDR) protocol of non-adaptive p-persistent CSMA is very far [24] along and appears to be compatible with the OSI model and ATN.

Even if a RA scheme is compatible with ATN, there is still the question of overhead in terms of meeting expected packet delay requirements. In other words, there are a fair number of extra "signal units", involving a considerable number of extra bytes (of 8 bits each) required to be compatible with ATN. Even if a basic RA scheme <u>might</u> meet a particular delay requirement without being compatible with ATN, the required delay performance <u>might not</u> be met with the overhead imposed by ATN. This issue needs to be addressed in evaluating options for the transition phase.

Another important aspect to consider for transition, as with any software driven functionality, is to ensure that any replacement radios have the capacity to handle software upgrades. Oftentimes the full software functionality is not available when new hardware capabilities are first introduced. Clearly, the life cycle of new airborne equipment can be extended if there is adequate space in the avionics box for additional microprocessors or memory boards. Packaging problems can also be eased if early plans are made for downsizing subsystem components of the radio considering newer layout technologies. It is then feasible to modernize or upgrade the radio in stages by replacing subsystem modules without increasing the form-fit dimensions. In this way additional software functionality, such as a new RA protocol, can be introduced by changing-out some cards in the processing "slice" of the radio, for example, without requiring any modifications to the aircraft's existing physical or electrical configurations.

If a RA protocol is to be used in a future radio system, its particular needs for memory and processing power (speed) should be estimated well in advance to assure that the software capabilities can evolve in a preplanned fashion without requiring major changes to the radio hardware, especially that associated with the analog intermediate frequency (IF) and radio frequency (RF) components. As may be surmised from this paper, none of the RA

protocols discussed should have a large impact on the radio hardware. For example, the absolute timing requirements of no better than 1 ms or so (see Table 11-4) are quite modest by today's technology standards. Thus, high quality frequency standards (oscillators) or elaborate timing protocols are unnecessary.

In summary, the use of RA protocols in an evolutionary series of VHF A/G communications system upgrades should not pose any serious problems. The software protocols should be ATN compatible and the processing architecture should be thought through in advance. This would facilitate the introduction of unforeseen changes during the transition period.

11.7 CRITERIA FOR DECISION

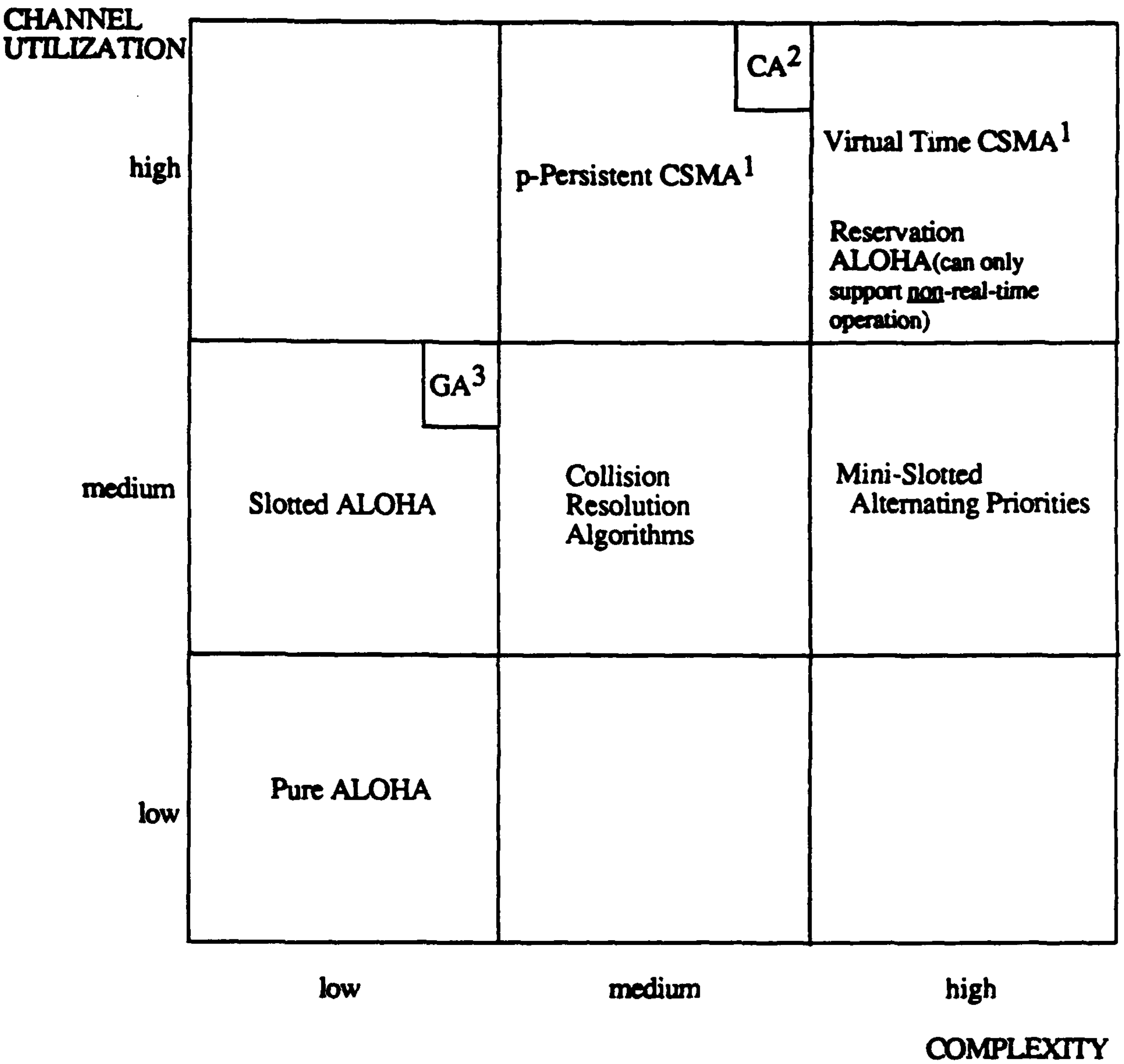
The main conclusion of the foregoing analysis might well be that none of the RA schemes discussed can be expected to support real-time (200 ms latency) operations at sufficiently high levels of channel utilization. Although it is true that all of these schemes but Reservation ALOHA can support near-real-time operations (1s latency), or even real-time operations if the messages are short enough, this must be tempered by the fact that only the expected (50%) delay measure has been used as the latency criteria. For acceptable real-time digital voice operation, for example, a more robust measure of packet delay, e.g., a 95% (or higher) level of confidence of receiving a transmitted packet correctly within 200 ms, should be applied, since some but not many digital voice packets can be "lost" before intelligibility is impaired. This conclusion appears to be well supported by other analyses [19, 23], as well. Furthermore, considering the maturity of the field, it is unlikely that any RA exists that could support real-time operation.

If one accepts that only near-real time operation with transmit-to-receive delays in the order of seconds is possible, the question is, which RA schemes are most effective? Within the context of aeronautical A/G communications we submit that channel throughput, or more generally channel utilization, and algorithm complexity are the two most important dimensions in the decision space. Based on the foregoing analysis, we have placed all seven RA schemes discussed, in this two-dimensional domain, as shown in figure 11-13. It may be helpful for the reader to refer back to Table 11-4 of the tradeoffs section and figure 11-12 of the impact/importance section.

Clearly one wishes for high throughput at low complexity, i.e., a RA scheme that would fit the upper-left-hand square of the 3×3 array of figure 11-13. Unfortunately, no such scheme appears to exist. We might be able to accept lower throughput with no increase in complexity, or higher complexity with no decrease in throughput. These alternatives correspond to Slotted ALOHA and p-persistent CSMA, respectively. As indicated, General Aviation (GA) users might be content with the throughput-delay performance of Slotted ALOHA provided they are attracted to the features of the radio in which this RA protocol is embedded. Presumably low complexity would translate into a low manufacturing cost and a price affordable by the GA community. On the other hand, Commercial Aviation (CA) users can afford to pay more to achieve the higher throughput that is urgently needed for AOC applications, in particular. Hence, the CA community appears to prefer the p-persistent CSMA scheme. Although Virtual Time CSMA performs somewhat better than p-persistent CSMA, the additional complexity may not be justified, and this algorithm may not be of sufficient practical interest in aeronautical applications.

The other RA schemes shown in figure 11-13 appear to be less desirable than Slotted ALOHA or p-persistent CSMA. For example, CRAs are superior to Slotted ALOHA with respect to stability, but not modified Slotted ALOHA, and CRAs are significantly more complex. The mini-slotted alternating priorities algorithm is relatively complex and only achieves medium channel utilization levels under rather heavy channel loading conditions. Pure ALOHA, although the simplest scheme has significantly lower throughput compared to Slotted ALOHA, where the timing accuracy necessary for Slotted ALOHA, and the other schemes for that matter, is easy to achieve in this application.

In summary, the preferred RA schemes seem to be Slotted ALOHA if a lower throughput is acceptable in meeting certain ATS and AOC requirements, and p-persistent CSMA when higher throughput is required and the level of complexity is acceptable. However, neither scheme is truly suitable for real-time operations such as real-time digital voice.



Notes:

1. Carrier Sense Multiple Access
2. Commercial Aviation selection
3. General Aviation possibility

Figure 11-13. Qualitative Assessment of Random Access Algorithms with Respect to Their Feasibility of Supporting Real-Time Operations

11.8 CONNECTIVITY/RELATIONSHIP WITH OTHER DECISIONS

The reader is referred to the decision tree structure depicted in appendix A. RA techniques are covered by branchpoint 4.2.3 under the Channel Access branchpoint in the Communications Throughput tree. RA schemes as a class are the principal alternatives to Fixed Assigned Channel Access schemes, branchpoint 4.2.1, which are inflexible, and Demand Assigned Channel Access schemes, branchpoint 4.2.2, which are more complex but can adapt dynamically to large variations in channel loading. Thus far, we have not generated a decision tree paper dealing specifically with Demand Assigned Channel Access. However, it is noted that RA schemes can be used to make requests for channel capacity within a Demand Assigned scheme. The Reservation ALOHA technique, briefly mentioned in the issues section, is an example that can be easily used with Demand Assigned FDMA or TDMA; see branchpoints 4.2.2.1 and 4.2.2.2.

RA techniques only apply to digital modulation and coding schemes (branchpoint 4.1.2). Thus, RA algorithms cannot be employed with analog modulations (branchpoint 4.1.1).

Similarly, the RA schemes emphasized in this paper are employed by users that are competing for the time resource on a single frequency channel. Thus, these principal RA techniques cannot be used with other than TDM Multiplexing schemes (branchpoint 4.3.2.2), or with Frequency Diversity schemes (branchpoints 2.1.1.1.1 and 2.2.1.4). However, RA can be used for interaircraft communication when the uplinks and downlinks are on the same frequency (branchpoint 2.5.1.1). RA might also help if one is trying to Minimize Frequency Changes (branchpoint 2.3.1.2). In addition, Message Coding (branchpoint 2.2.1.2), ARQ (branchpoint 2.2.1.3), and Equalization (branchpoint 2.2.1.5) can be used with RA techniques.

As we have seen RA techniques cannot be used with confidence for our real-time digital voice applications (branchpoint 2.2.1.1.2). Even more importantly, RA schemes cannot easily be used for Busy Channel Handling (branchpoint 2.1.3), especially Emergency "Break-in" (branchpoint 2.1.3.1) where a very quick response is desired, and where a more "disciplined" time structure involving more than a single frequency channel is preferred.

The simpler RA schemes like ALOHA would contribute to greater Robustness (branchpoint 2.1.2). If increased Frequency Stability (branchpoint 6.1.1) were attained under a Closer Channel Spacing option, this might benefit the slotted RA schemes (all but Pure ALOHA), but an accurate time reference is really not required.

RA techniques are potentially appropriate for many near-real time or non-real time data link applications (branchpoint 2.4). AVPAC (branchpoint 2.4.5) is the most relevant case in point in our discussions. Here p-persistent CSMA (branchpoint 4.2.3.1.2.3) has been selected as the Channel Access scheme, and an increased data rate has been emphasized as a way of increasing channel capacity and reducing access delay through (branchpoint 2.2.2) through shorter packets and fewer packet collisions.

LIST OF REFERENCES

References for Section 5

1. Goodman, D. J., *Trends in Cellular and Cordless Communications,* IEEE Communications Magazine, June 1991, pp 31-40.

2. Noguchi, T., Y. Daido, and J. A. Nossek, *Modulation Techniques for Digital Microwave Radio,* IEEE Communications Magazine, October 1986, pp 21-30..

References for Section 6

1. Swartz, M., W. R. Bennett, and S. Stein, *Communication Systems and Techniques,* McGraw Hill, New York, 1966, pp 120-121.

2. Special Issue on Advanced Mobile Phone Service, The Bell System Technical Journal, Vol. 58, No. 1, January 1979.

3. Proakis, J. G., *Digital Communications*, McGraw Hill, New York, 1983, pp 193-199.

References for Section 7

1. Shannon, C. E., *A Mathematical Theory of Communication,* Bell System Technical Journal, volume 27, pp 623-656, October 1948.

2. Hayking, S., *Digital Communications,* Wiley, New York, 1988.

3 Callaghan, T.G., J. L. Perry and J. Tjho; *Modulation Recognition for Automated Surveillance Systems*; IEEE Transactions on Communications, Oct. 1985.

References for Section 8

1. Proakis, J. G., *Digital Communications*, McGraw-Hill Inc., New York, NY, 1983, p. 148.

2. Wachira, M., H. Yazdani, K. Feher, and W. Steencart, *A Survey of Eight-Phase PSK and QPRS Digital Radio Systems*, Proc. 1978 IEEE Canadian Communications and Power Conference, Montreal, October 1978.

3. *NEC Digital Radio System (11 GHz 1344-CH) System Application Manual,* MSD-3902; 7509-02, Nippon Electric Co., Ltd., Tokyo, 1975.

4. Morais, D., and K. Feher, *A Survey of North American 2 GHz Digital Radio Systems*, Proc. Intelcom-77, Atlanta, October 1977.

5. Pasupathy, S., *Minimum Shift Keying: A Spectral Efficient Modulation,* IEEE Communications Magazine, July 1979, pp. 14-22.

6. Murta, K., and K. Hirade, *GMSK Modulation for Digital Mobile Radio Telephony*, IEEE Trans. Commun., Vol. COM-29, No. 7, July 1981, pp. 1044-1050.

7. Casewell, I. E., *The Provision of GSM Cellular Radio Environment within Passenger Aircraft Operating over Europe*, IEE Fifth International Conference on Mobile Radio and Personal Communications, No. 315, December 1989, pp. 172-176.

8. Akaiwa, Y. and Y. Nagata, *Highly Efficient Digital Mobile Communication with a Linear Modulation Method*, IEEE Trans. on Selected Areas in Commun., Vol. SAC-5, No. 5, June 1987, pp. 890-895.

9. Electronics Industries Association, *Cellular System Dual-Mode Mobile Station - Base Station Compatibility Standard*, EIA/TIA Interim Standard, IS-54, May 1990.

10. AEEC Project Paper 750, The ARINC Companies, October 25, 1991.

11. Feher, K., *A New Generation of 90 Mb/s SATCOM System: Bandwidth efficient, field tested 16-QAM*, ICC-89, Boston, MA, Vol. 1, pp. 18.1.1-18.1.6.

12. Nossek, J. A., and H. Barth, *16-State QAM-Modem for 140 Mb/s Digital Radio System Family*, IEEE Globecom '83, San Diego, CA, Conf. Rec., Vol. 1, pp. 9.1.1-9.1.6.

13. Lucky, R. W., J. Salz, and E. J. Weldon, Jr., *Principles of Data Communications*, McGraw-Hill Inc., New York, NY, 1968, p. 202.

14. Amoroso, F., *The Bandwidth of Digital Data Signals*, IEEE Communications Magazine, November, 1980, pp. 13-24.

15. Korn, I., *Digital Communications*, Van Norstrand Reinhold Co., New York, NY, 1985, p. 432.

16. Wilson, W. J., *Air-Ground Communications Options for Air Traffic Control*, M92B0000017, The MITRE Corporation, February 1992.

References for Section 9

1. Peterson, W. W. and E. J. Weldon Jr., *Error-Correcting Codes*, 2nd ed., MIT Press, Cambridge, MA, 1972.

2. AEEC Project Paper 750, The ARINC Companies, October 25, 1991.

3. McWilliams, F. J. and J. A. Sloane, *The Theory of Error Correcting Codes*, North-Holland, Amsterdam, 1977, p. 298.

4. Proakis, J. G., *Digital Communication*, McGraw-Hill, New York, NY, 1983, p. 186.

5. Skar, B., *Digital Communications Fundamental and Applications*, Prentice-Hall, Englewood Cliff, NJ, 1988, p. 306.

6. Ibid, p. 307.

7. Chen, C-H., *Channel Bit-Error-Rate versus Packet Error Rate*, Memo D053-M-1539, The MITRE Corporation, 9 December 1991.

8. Lin, S., and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1983, pp. 458-494.

References for Section 10

1. Martin, J., *Data Communications Technology*, Prentice Hall, Englewood Cliffs, New Jersey, 1988, p. 146.

2. Freeman, R. L., *Telecommunication Transmission Handbook*, John Wiley & Sons, New York, 1981, p. 135.

3. Fluhr, Z. C., and P. T. Porter, *Control Architecture*, Bell System Technical Journal, 58, No. 1, January 1979, pp. 43-69.

4. Wilson, W. J., *Solution to the CTAG Cellular Problem*, MITRE Memo D053-M-024, 23 May 1991.

5. Wilson, W. J., *CTAG Frequency Reuse Patterns and Propagation Effects*, MITRE Memo D053-M-1454, 15 October 1991.

6. Elnoubi, S. M., *Coverage and Interference Considerations in the 3-Dimentional Cellular Plan*, MITRE Memo F085-M-730, 16 December 1991.

7. Proakis, J. G., *Digital Communications*, McGraw-Hill, New York, 1983, pp. 171-178.

References for Section 11

1. Rucker, R. A., ed., *Program Plan for the ATN Project*, ML-91W00008, The MITRE Corporation, McLean, VA 22102, May 1991.

2. Brennan, V., *Draft Requirements for the Future VHF Air-Ground Radiotelephony System*, Working Paper (WP/66), RTCA, Special Committee 172, Working Group 1, 6-8 June 1992.

3. *NAS System Requirements Specification*, NAS-SR-1000, Federal Aviation Administration, 21 April 1989.

4. *Minutes of Third Meeting*, January 13-14, 1992 RTCA SC-172 WG-1 WP/39, VHF Air-Ground System Improvements Study, RTCA SC-172, Working Group 1, March 26, 1992, pp. 12-14.

5. *Third Annual International Aeronautical Telecommunications Symposium on Data Link Integration*, Proceedings, McLean, VA, 20-23 May 1991.

6. *Signal-In-Space Minimum Aviation System Performance Standards for Advanced VHF Digital Data Communications Including Compatibility with Digital Voice Techniques*, Paper 682-92/SC172-58, RTCA, Sixth Draft, 28 September 1992.

7. *Aircraft Communications Addressing and Reporting System (ACARS)*, ARINC Characteristic 724, Aeronautical Radio, Inc.

8. Hanna, S., B. Mazur and M. Moyer, *Comparison of Access Strategies for VHF Digital Radio in ATC Applications*, Information Paper, AMCP/1-WG-C/1-WP [?], Agenda Item No. 5: Implementation Issues, Aeronautical Mobile Communications Panel (AMCP), Working Group-C Meeting, Kobe, Japan, 9-20 March 1992, p.8.

9. Tobagi, F. A., *Multi-access Protocols in Packet Communication Systems*, IEEE Trans. Commun., Vol. COM-28, No. 4, April 1980, pp. 468-488.

10. Kleinrock, L., and F. A. Tobagi, *Packet Switching in Radio Channels: Part I - Carrier Sense Multiple Access Modes and Their Throughput-Delay Characteristics*, IEEE Trans. Commun., Vol. COM-23, No. 12, December, 1975, pp. 1400-1416.

11. Lam, S. S., *Satellite Multi-Access Schemes for Data Traffic*, Proc. Int. Conf. Commun., Chicago, IL, 1977, pp. 37.1-19 to 37.1-24.

12. Massey, J. L., *Collision-Resolution Algorithms and Random-Access Communications*, in Multi-User Communications, G. Longo, Ed., CISM Courses and Lectures Series, No. 265, New York: Springer-Verlag, 1981, pp. 73-137.

13. White, B. E., *An Extension of Collision Resolution (Tree) Algorithms*, Conference Record, Vol. 1, IEEE Intern. Conf. on Commun., Boston, MA, 19-22 June 1983, pp. 392 - 396.

14. Kleinrock, L., and M. Scholl, *Packet Switching in Radio Channels: New Conflict-Free Multiple Access Schemes for a Small Number of Data Users*, ICC Conf. Proc., Chicago, IL, June 1977, pp. 22.1-105 to 22.1-111.

15. Molle, M. L., and L. Kleinrock, *Virtual Time CSMA: Why Two Clocks are Better Than One*, IEEE Trans. Commun., Vol. COM-33, No. 9, September 1985, pp. 919-933.

16. *Ibid.*, pp. 927-929.

17. Schwartz, M., Telecommunication Networks: Protocols, Modeling and Analysis, Addison-Wesley, Reading, MA, 1987.

18. Molle, M. L., *Unifications and Extensions of the Multiple Access Communications Problem*, Ph.D. Dissertation, Dept. of Computer Science, Univ. of California, Los Angeles, CSD Rep. 810730 (UCLA-ENG-8118), July 1981.

19. D'Amours, C., and B. A. Mazur, *Throughput and Delay Characteristics of AVPAC Radio*, Calian Communications Systems, Working Group (WG)-1, Working Paper (WP)/117; WG-2, WP/41, RTCA Special Committee (SC)-172, presented by Dave Sim, Transport Canada, 22 October 1992.

20. Friedman, S., personal communications, WG-2, RTCA SC-172, 1992.

21. Wilson, W. J., *Air-Ground Communications Options for Air Traffic Control*, M92B0000017, The MITRE Corporation, Bedford, MA, February 1992, pp. 7,8.

22. Elnoubi, S., and C. Moody, *Evaluation of Single Frequency Versus Dual Frequency Channel Structure for VHF A/G Communications*, WP92W0000334, The MITRE Corporation, McLean, VA, December 1992.

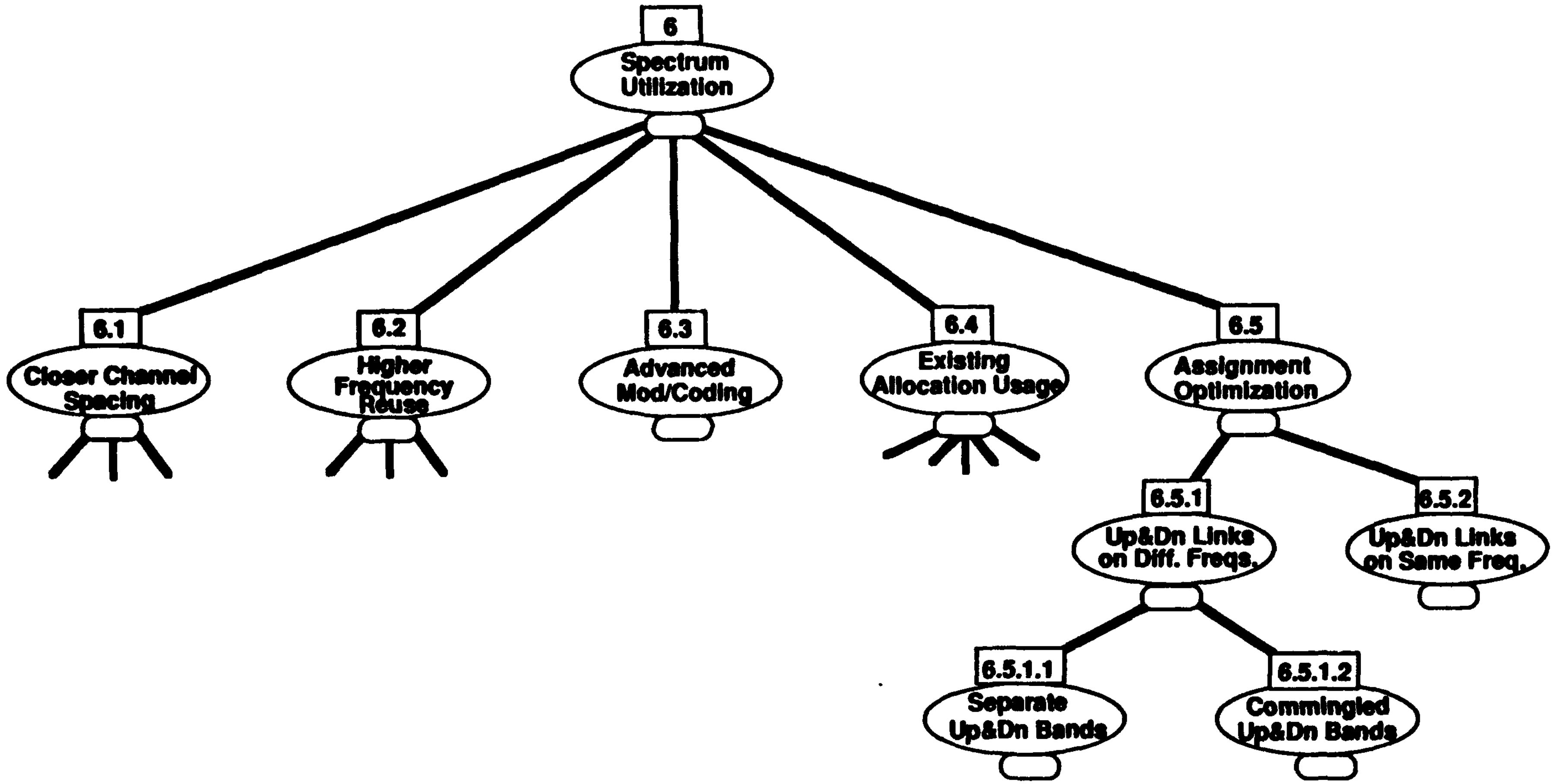
23. D'Amours, C., and B. A. Mazur, *Attachment to AVPAC Analysis*, Calian Communications Systems, Appendix to AMCP/2-WP/40, Agenda Item No. 8: Develop[ment] Material on VHF Digital Link (VDL), Second Meeting, Aeronautical Mobile Communications Panel (AMCP), Montreal, Canada, 23 November 1992, pp. 21-23.

24. *Draft 3 of Project Paper 750 VHF Data Radio*, Aeronautical Radio, Inc., Annapolis, MD, 14 September 1992.

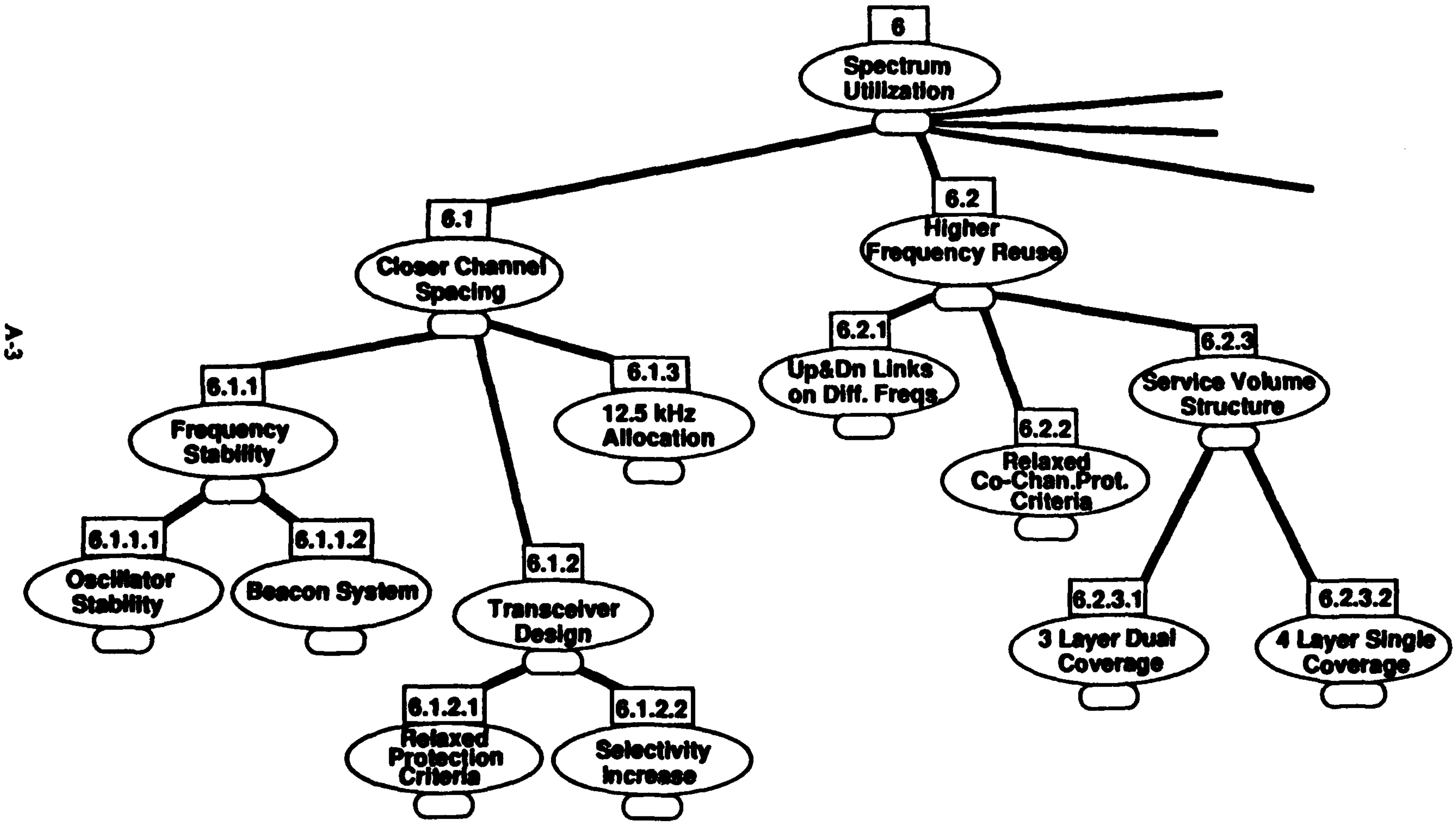
APPENDIX A

DECISION TREE SUBTREES

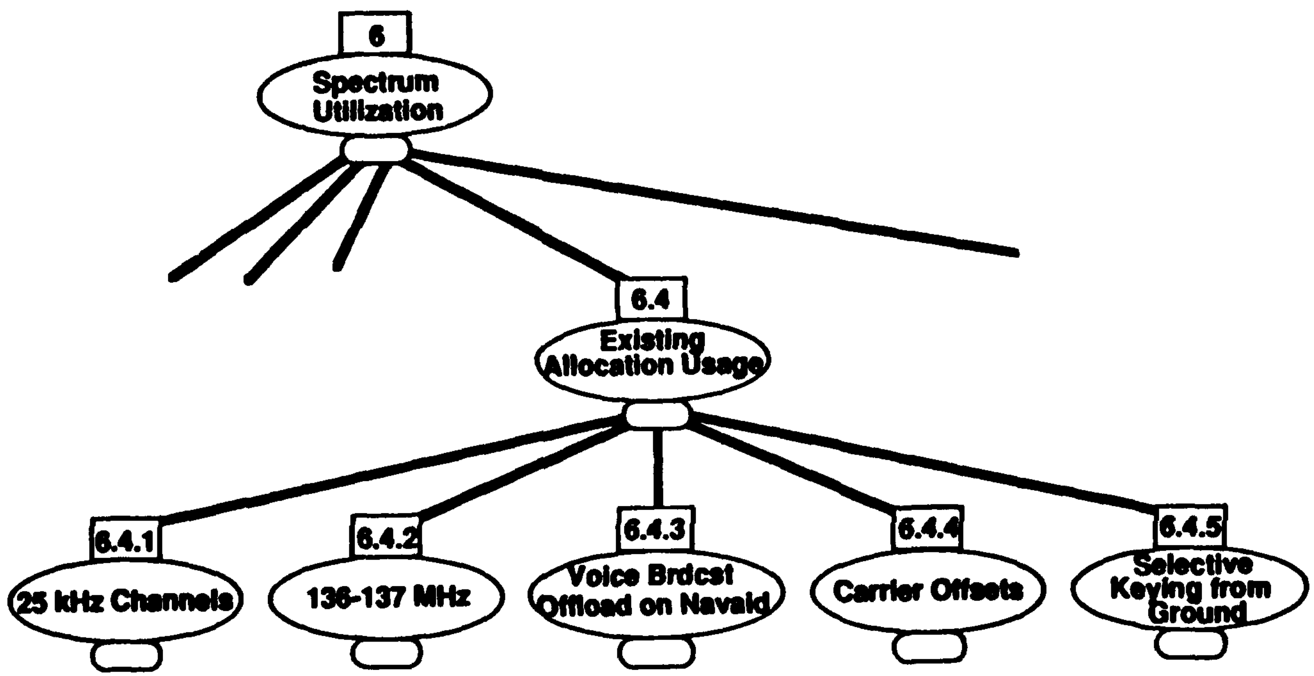
Spectrum Utilization Subtree (Top Levels) and Full Assignment Optimization Subtree



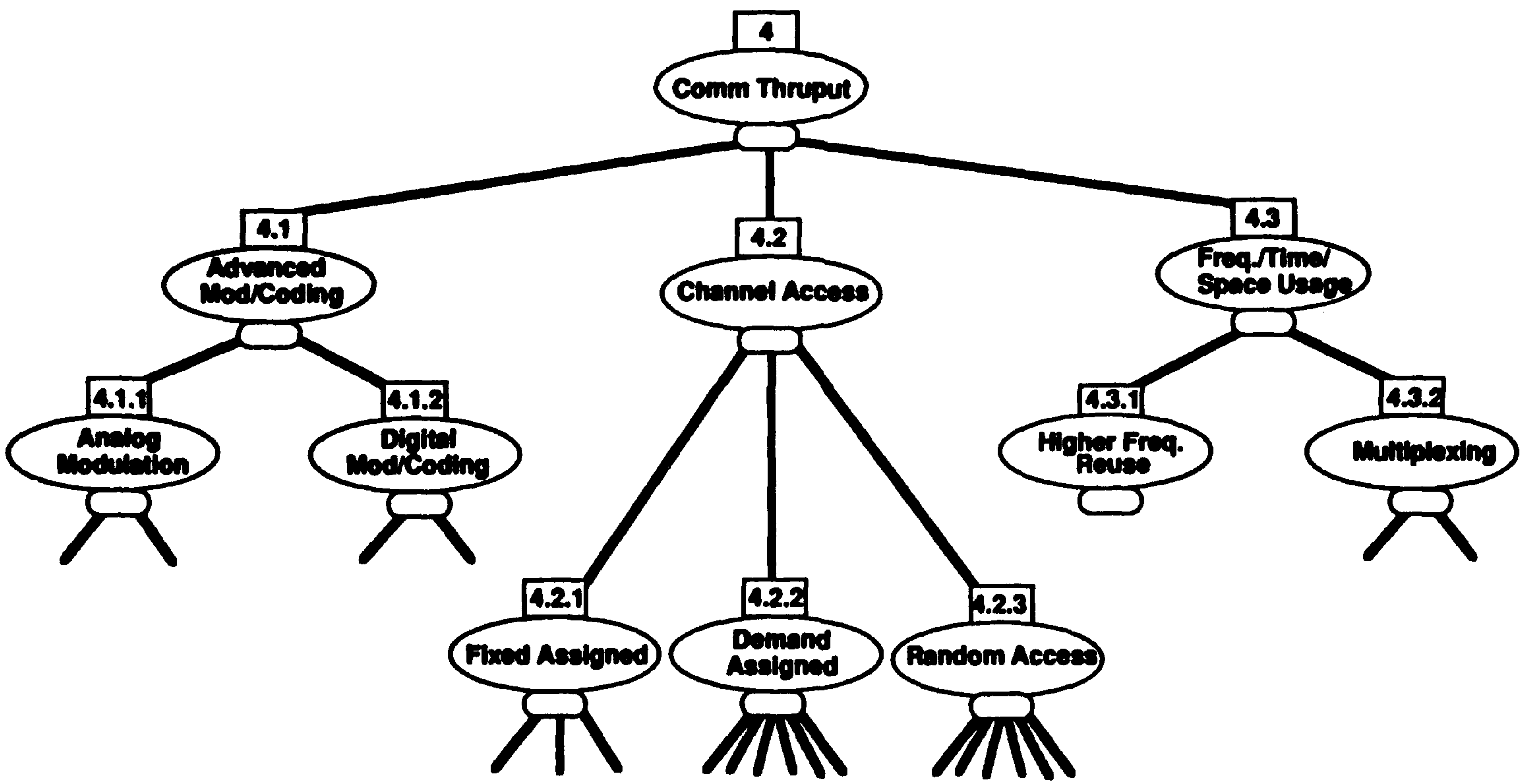
Full Closer Channel Spacing and Higher Frequency Reuse Subtrees



Full Existing Allocation Usage Subtree



Communications Throughput Subtree
(Top Levels)



Full Analog Modulation Subtree

AM - Amplitude Modulation

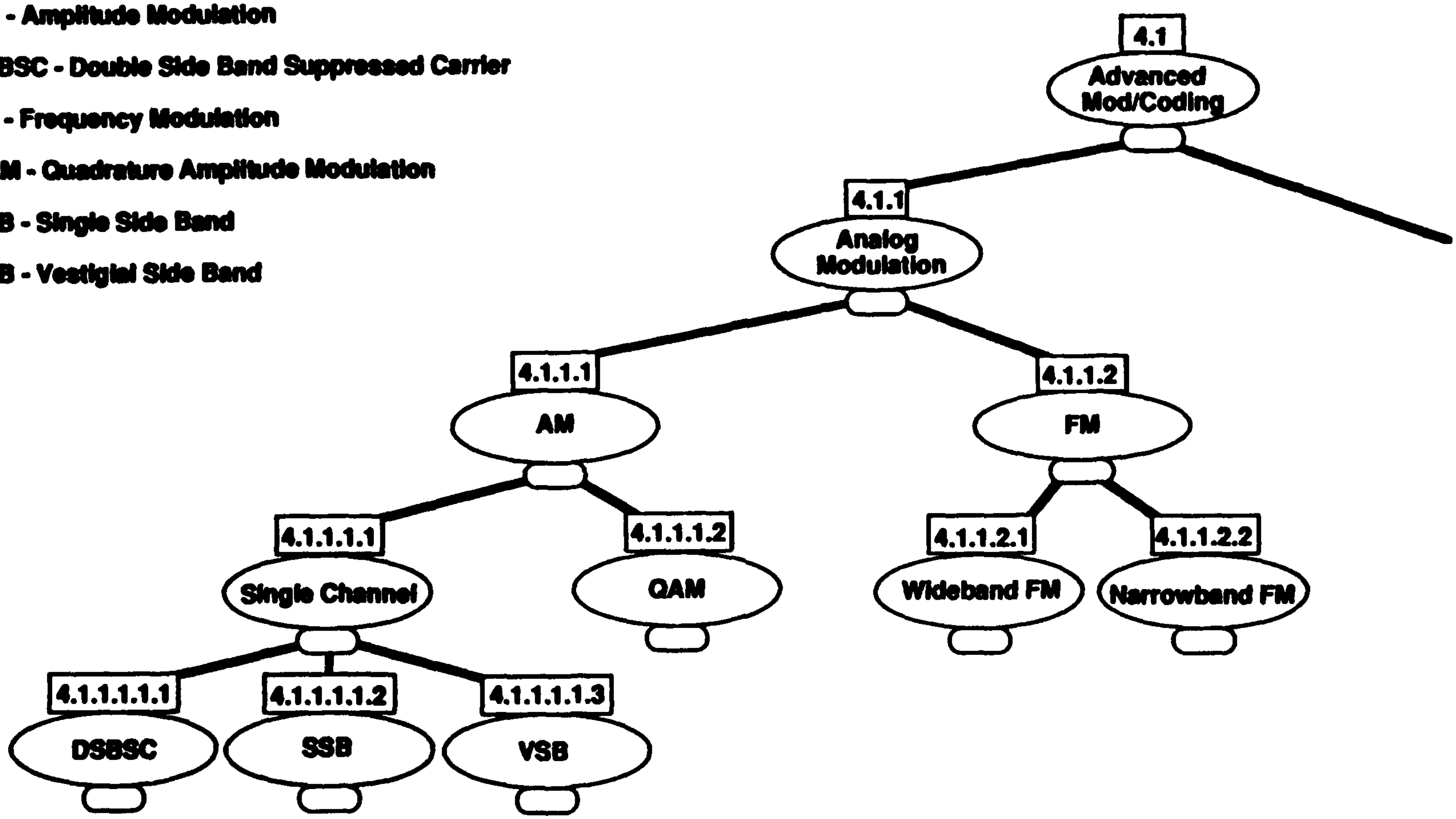
DSBSC - Double Side Band Suppressed Carrier

FM - Frequency Modulation

QAM - Quadrature Amplitude Modulation

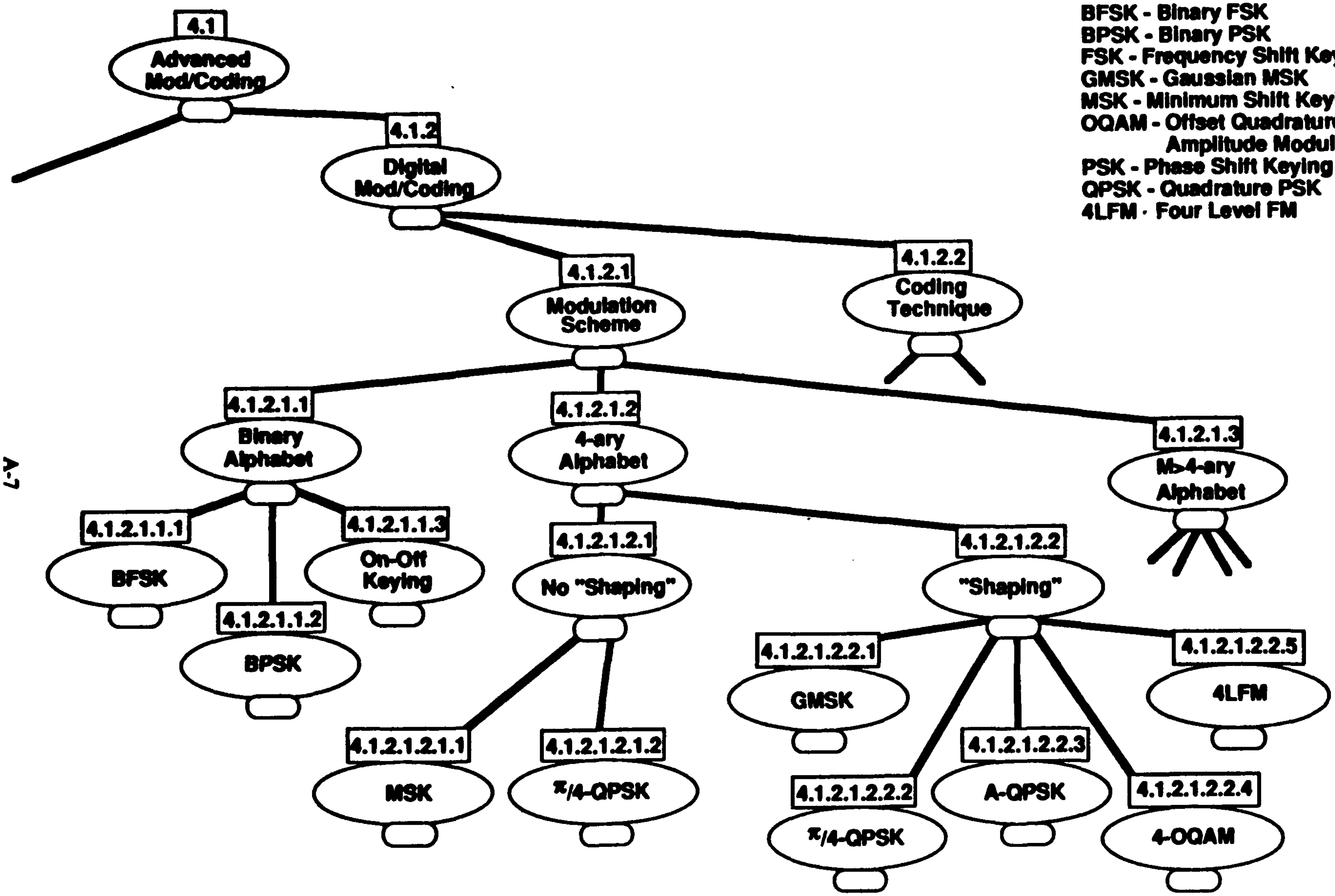
SSB - Single Side Band

VSB - Vestigial Side Band



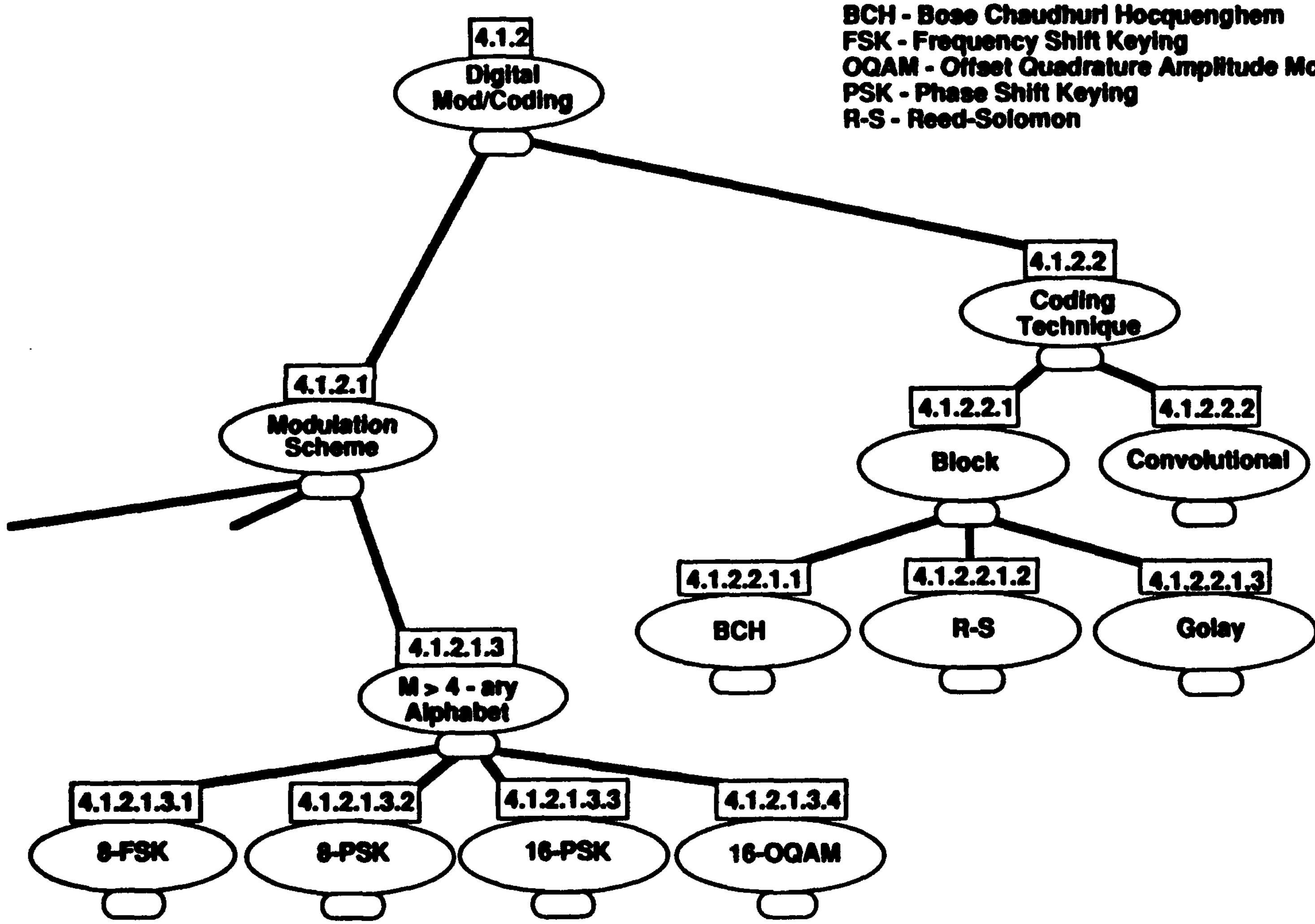
Partial Digital Modulation and Coding Subtree

A-QPSK - Aviation QPSK
BFSK - Binary FSK
BPSK - Binary PSK
FSK - Frequency Shift Keying
GMSK - Gaussian MSK
MSK - Minimum Shift Keying
OQAM - Offset Quadrature Amplitude Modulation
PSK - Phase Shift Keying
QPSK - Quadrature PSK
4LFM - Four Level FM



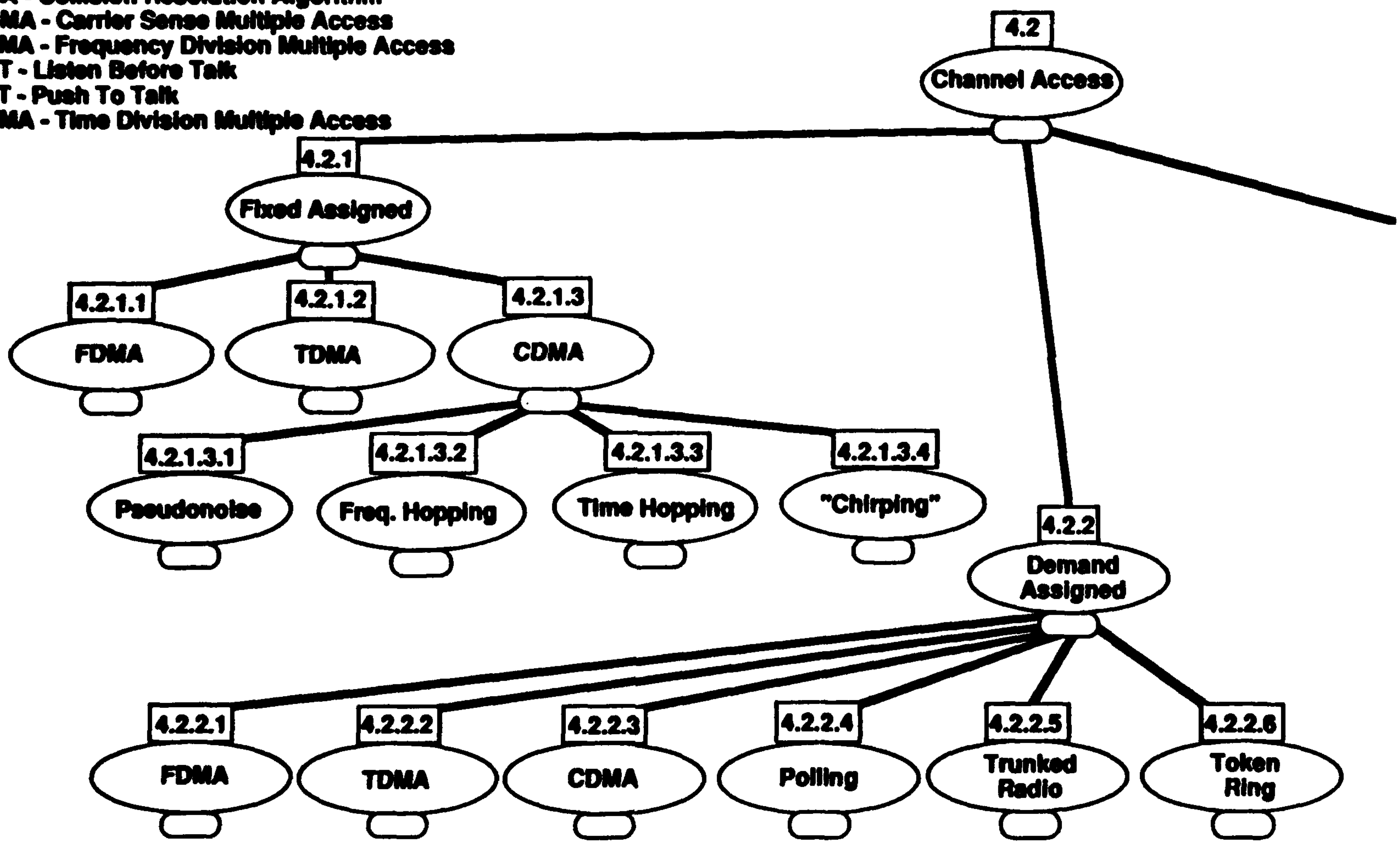
Full Coding Technique and M>4-ary Alphabet Subtrees

BCH - Bose Chaudhuri Hocquenghem
FSK - Frequency Shift Keying
OQAM - Offset Quadrature Amplitude Modulation
PSK - Phase Shift Keying
R-S - Reed-Solomon

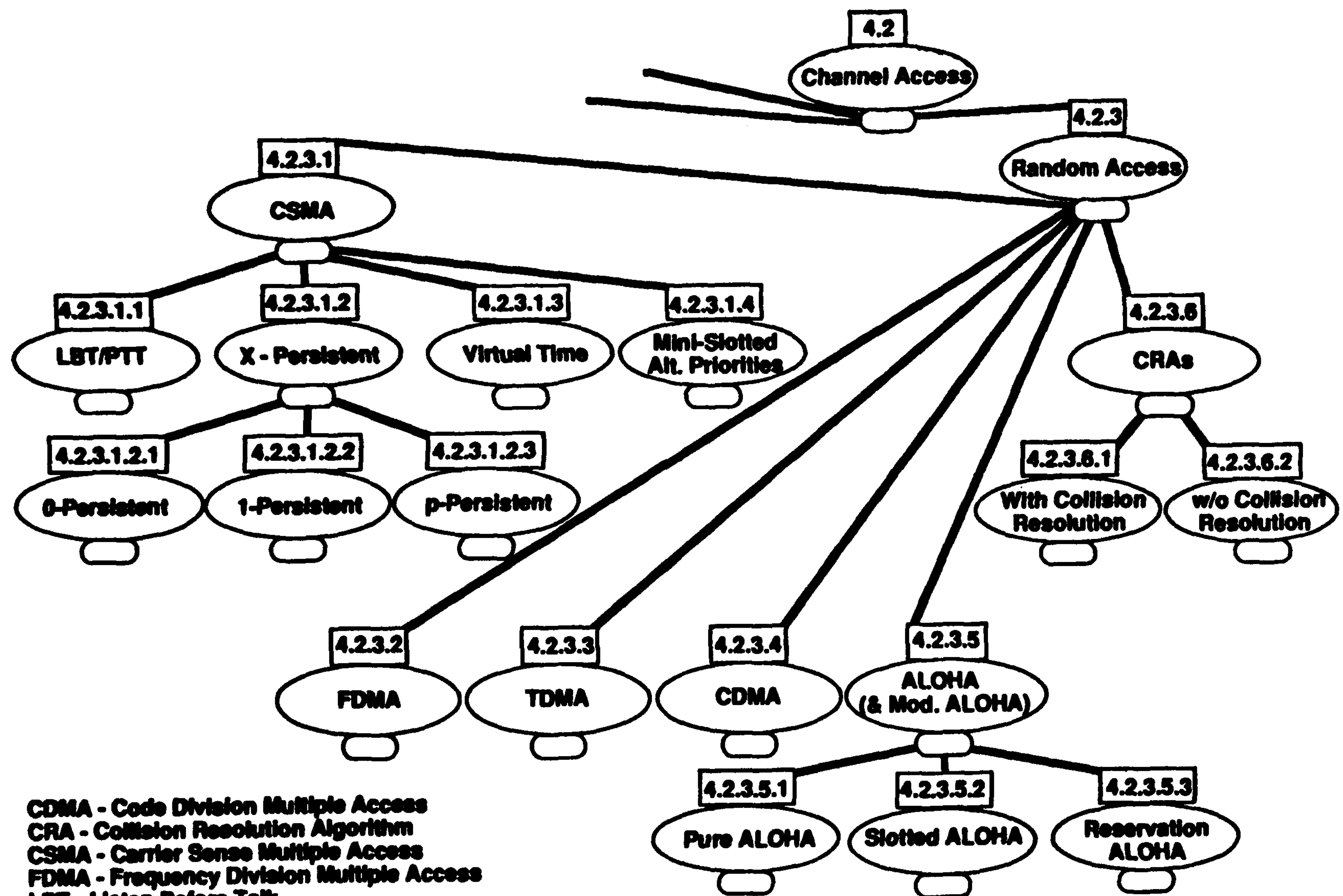


Full Fixed Assigned and Demand Assigned Subtrees

CDMA - Code Division Multiple Access
CRA - Collision Resolution Algorithm
CSMA - Carrier Sense Multiple Access
FDMA - Frequency Division Multiple Access
LBT - Listen Before Talk
PTT - Push To Talk
TDMA - Time Division Multiple Access

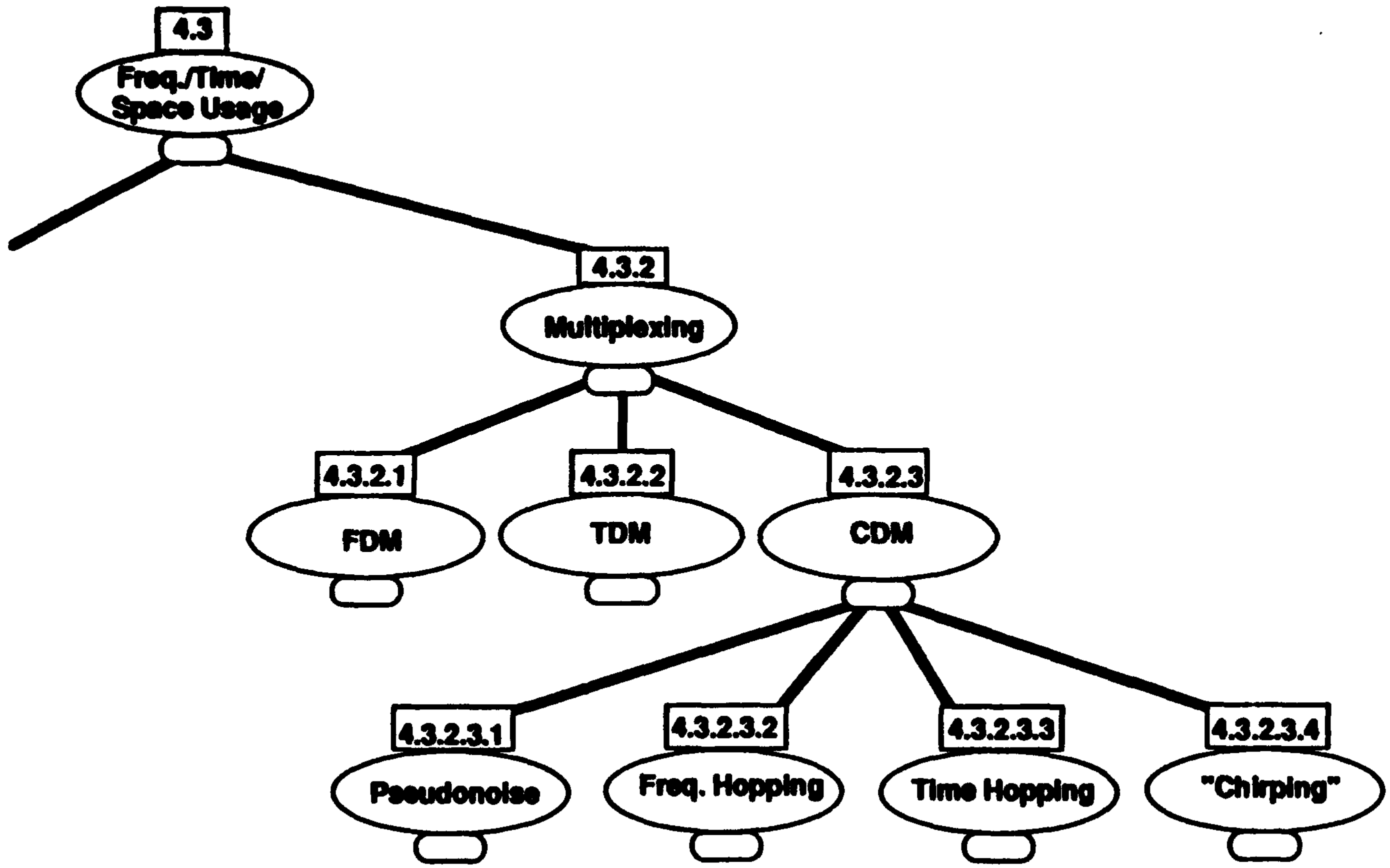


Full Random Access Subtree



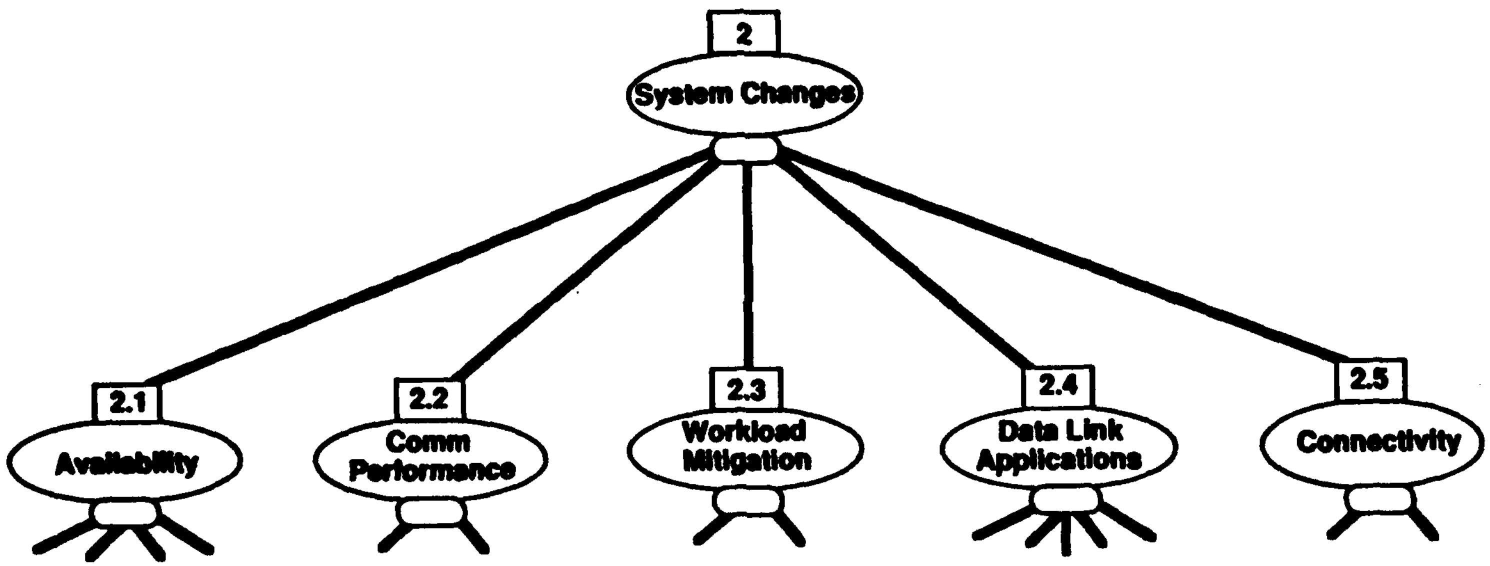
CDMA - Code Division Multiple Access
CRA - Collision Resolution Algorithm
CSMA - Carrier Sense Multiple Access
FDMA - Frequency Division Multiple Access
LBT - Listen Before Talk
PTT - Push To Talk
TDMA - Time Division Multiple Access

Full Multiplexing Subtree



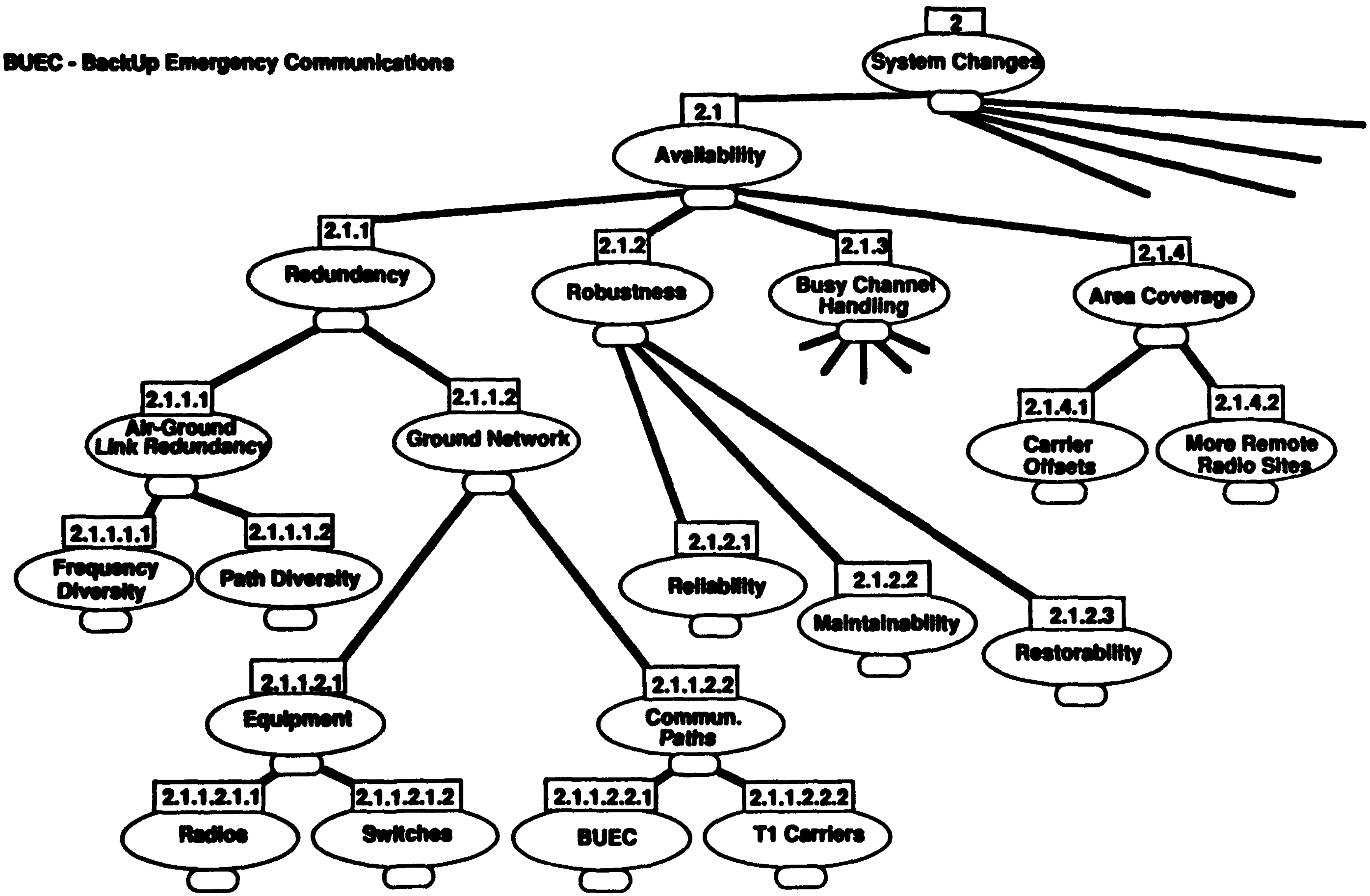
CDM - Code Division Multiplexing
FDM - Frequency Division Multiplexing
TDM - Time Division Multiplexing

System Changes Subtree (Top Levels)

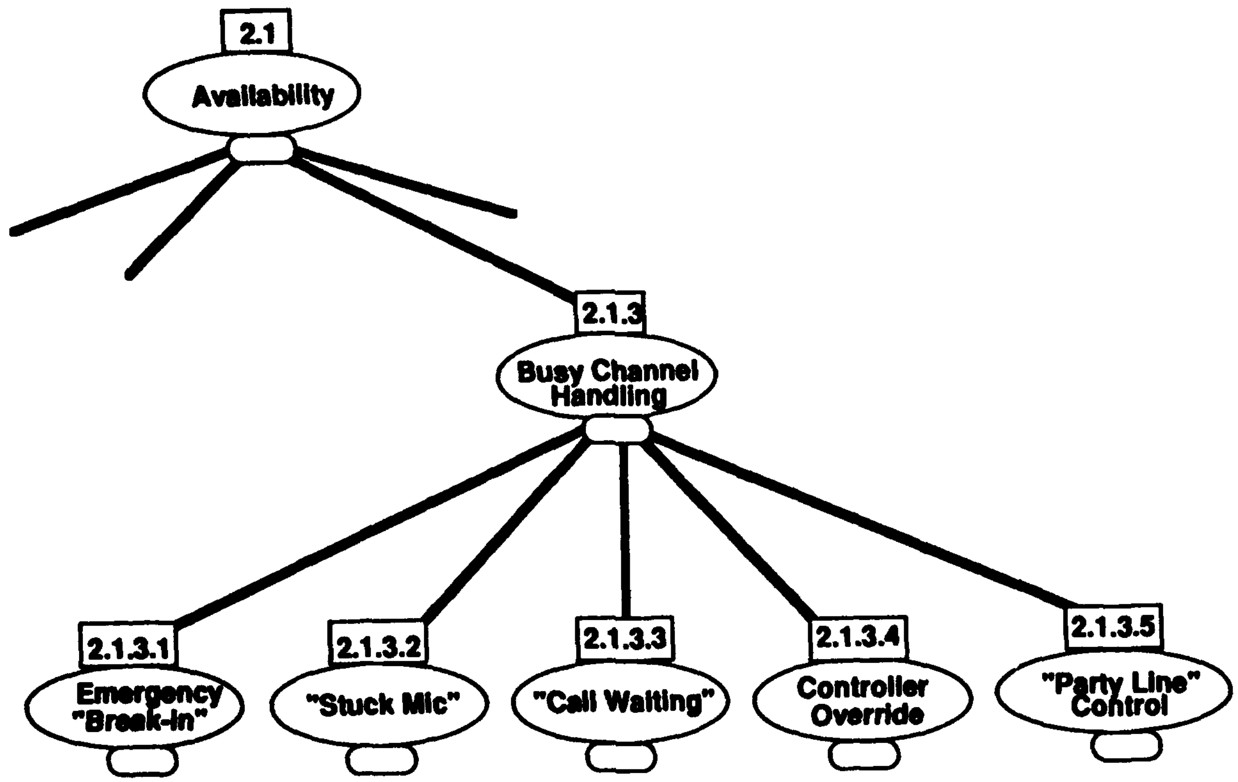


Partial Availability Subtree

BUEC - BackUp Emergency Communications

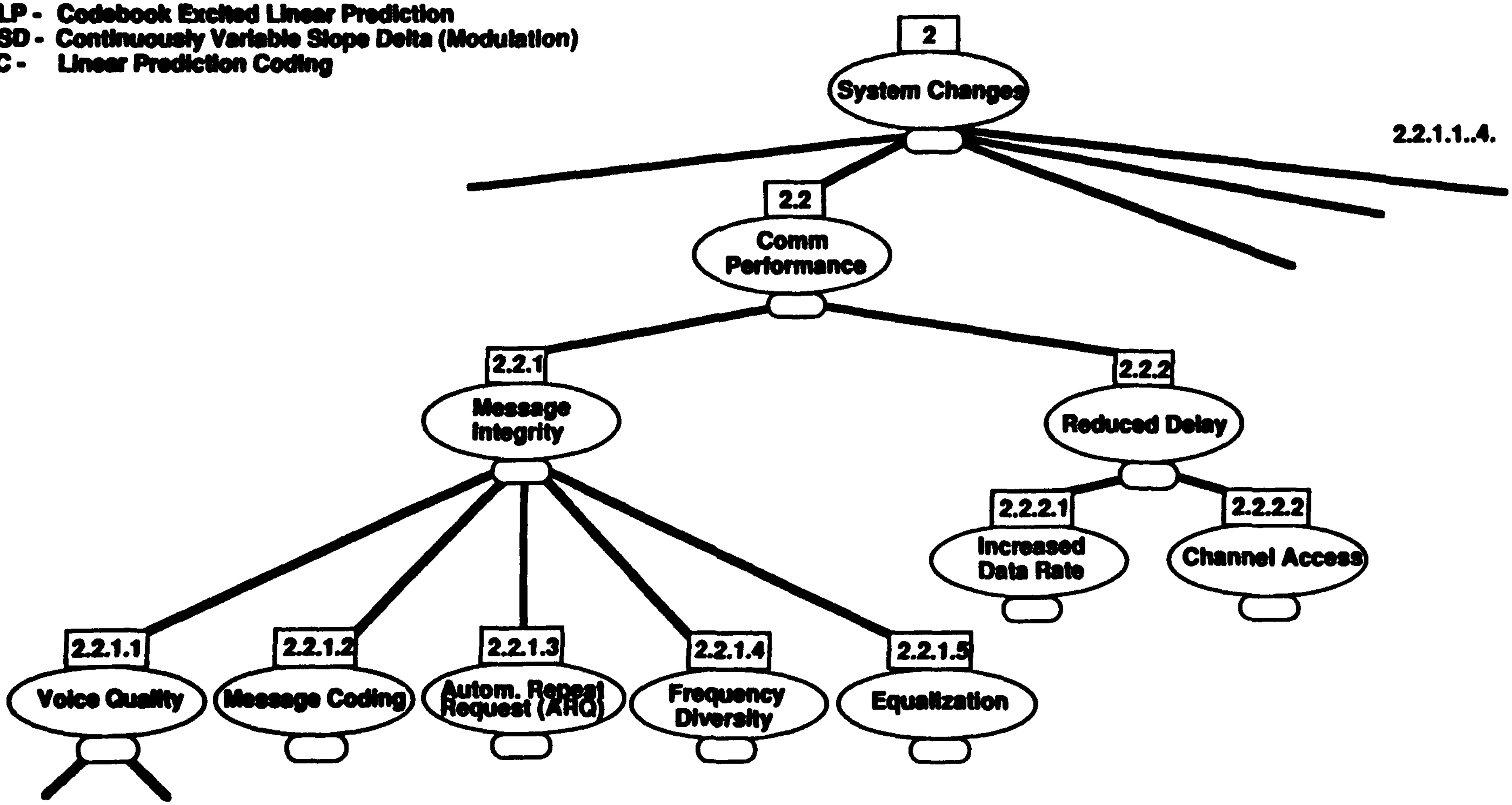


Full Busy Channel Handling Subtree



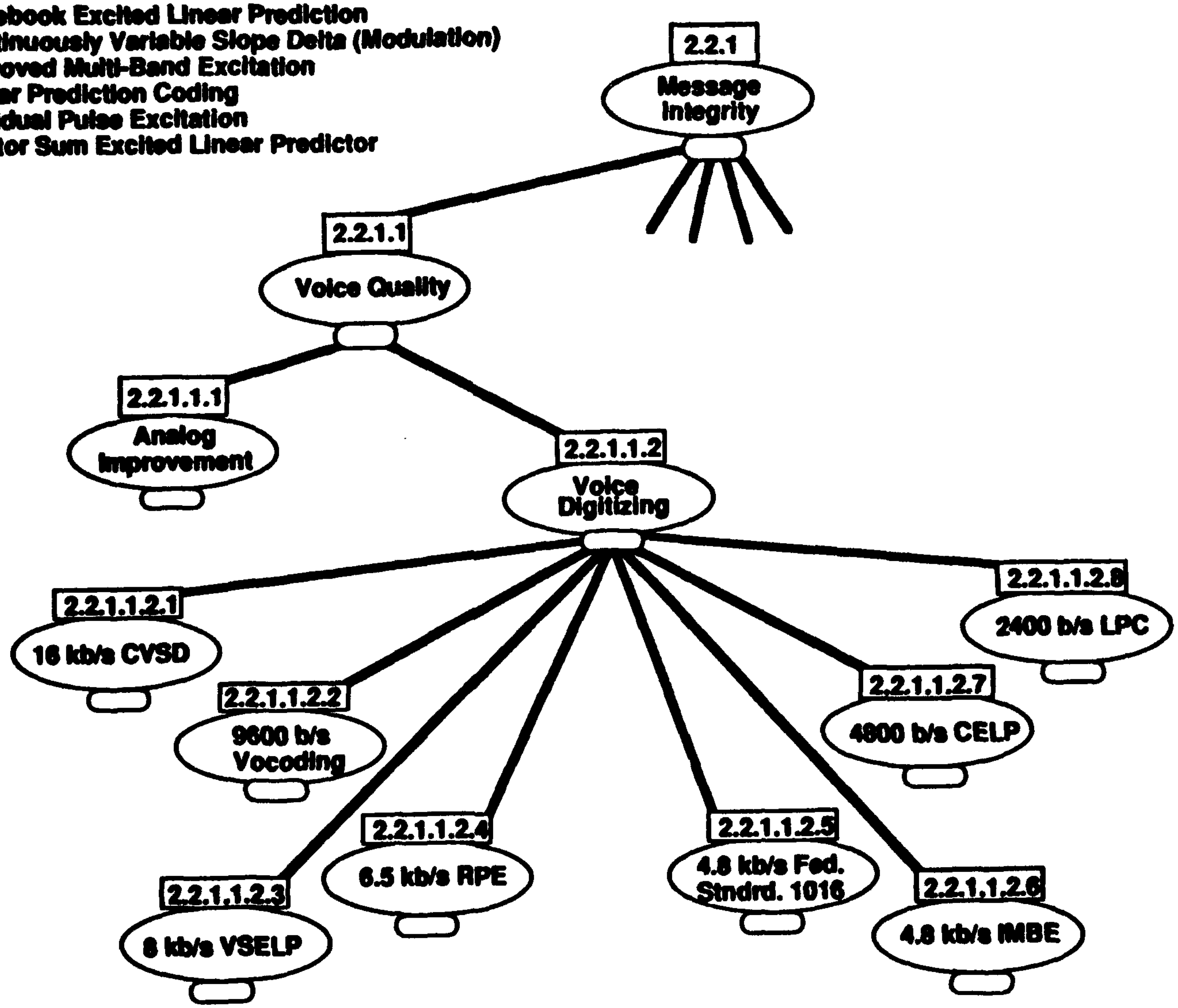
Partial Communications Performance Subtree

CELP - Codebook Excited Linear Prediction
CVSD - Continuously Variable Slope Delta (Modulation)
LPC - Linear Prediction Coding

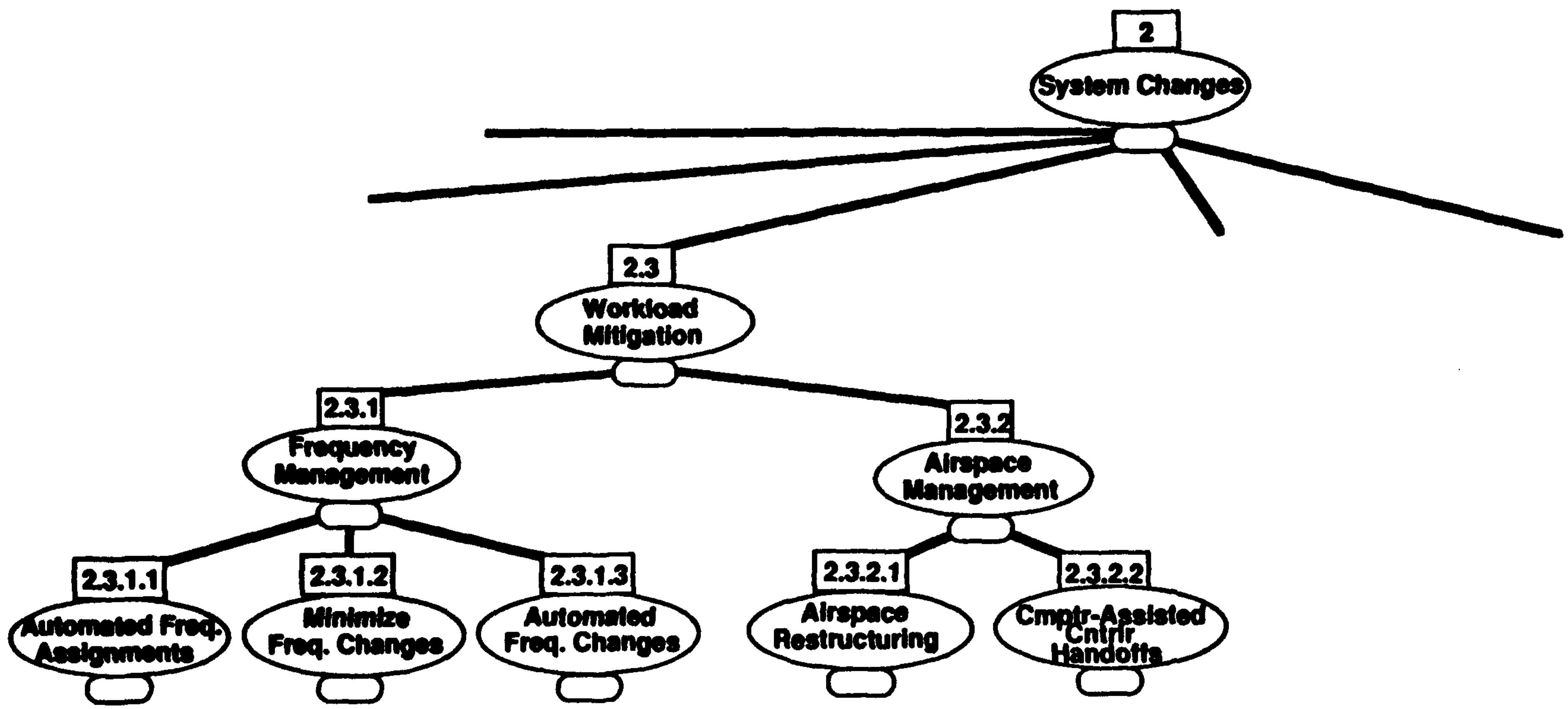


Full Voice Quality Subtree

CELP - Codebook Excited Linear Prediction
CVSD - Continuously Variable Slope Delta (Modulation)
IMBE - Improved Multi-Band Excitation
LPC - Linear Prediction Coding
RPE - Residual Pulse Excitation
VSELP - Vector Sum Excited Linear Predictor

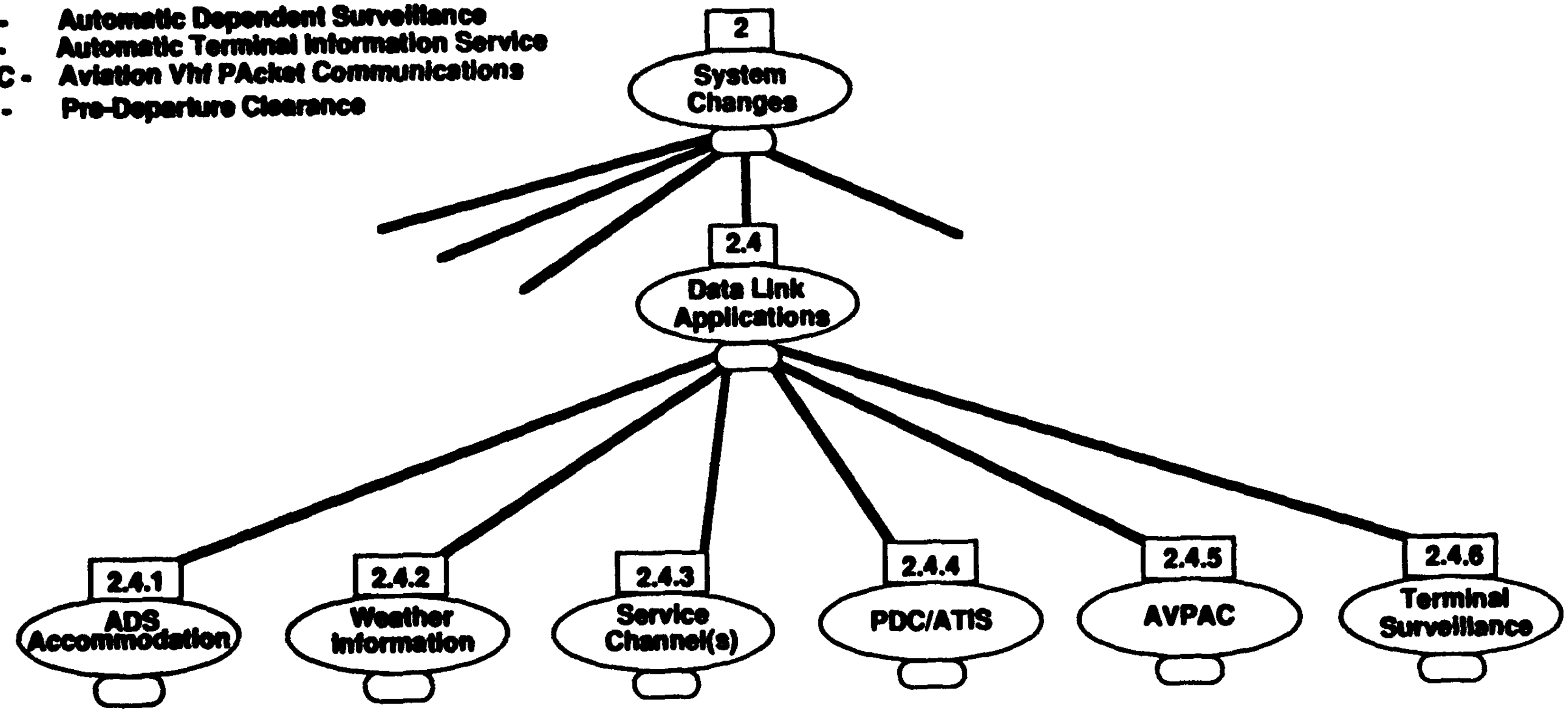


Full Decreased Workloads Subtree

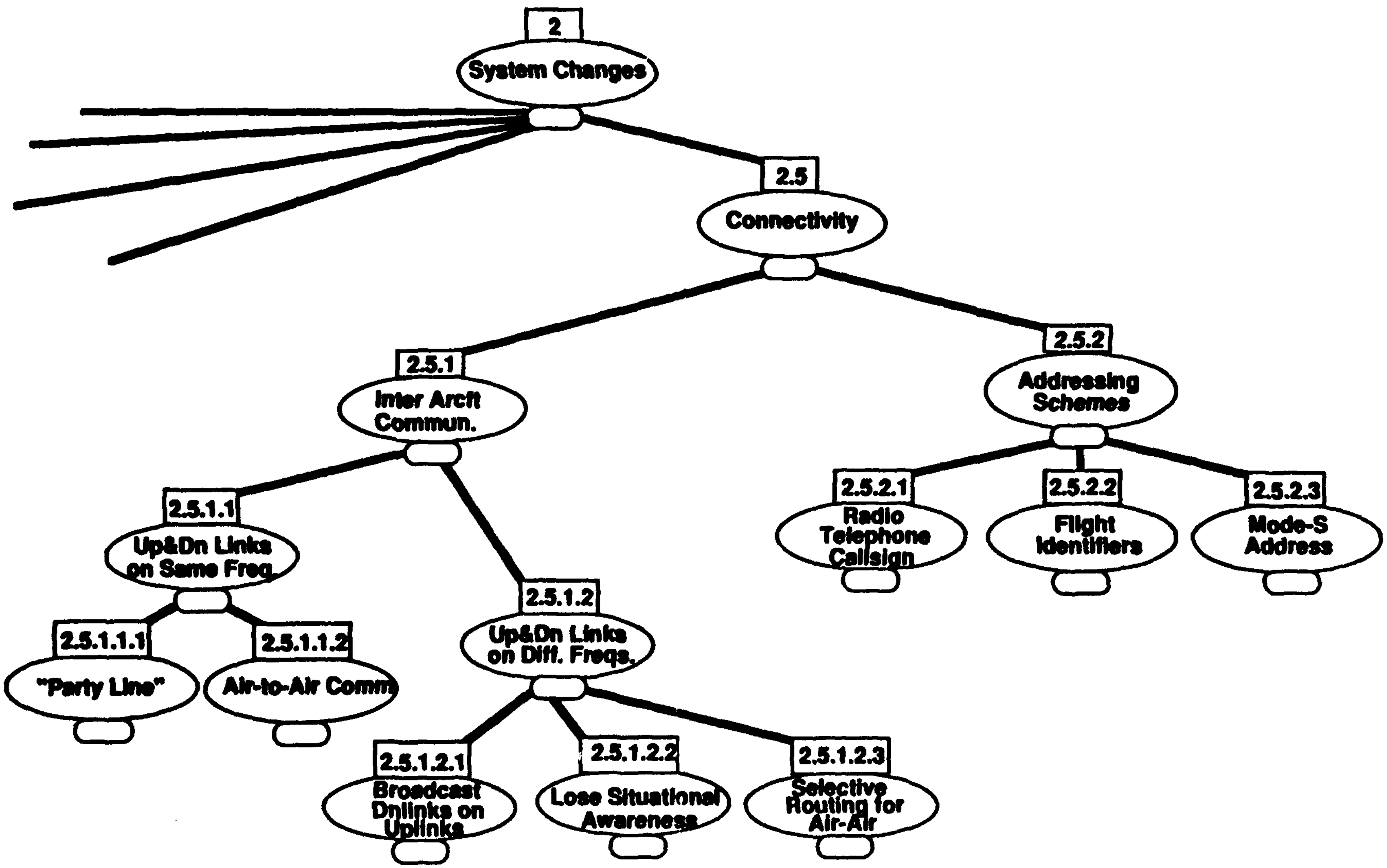


Full Data Link Applications Subtree

ADS - Automatic Dependent Surveillance
ATIS - Automatic Terminal Information Service
AVPAC - Aviation Vhf PAcket Communications
PDC - Pre-Departure Clearance



Full Connectivity Subtree



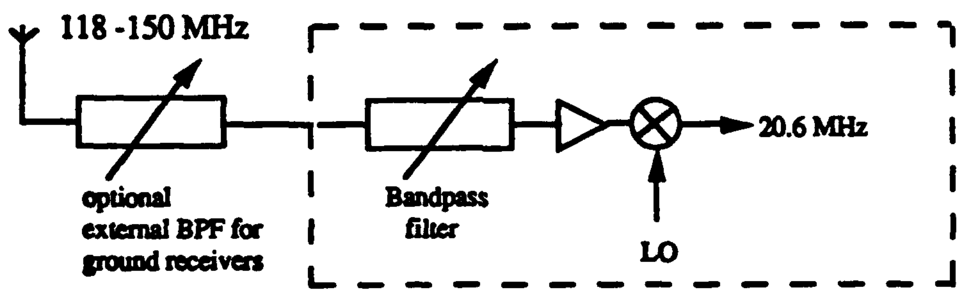
APPENDIX B

CO-LOCATED GROUND RADIO ISSUES

Multiple transmitters and receivers at a common site has been a difficult integration issue for many communication systems. Characteristics of both the transmitters and the receivers determine the degree of mutual and external interference. Strong signals close *and* far in frequency from the desired frequency can create additional frequency products within the passband of the victim receiver. Strong signals that fall within the RF bandwidth of a receiver close to the desired signal can *desense* the receiver or effectively degrade the ability of the receiver to detect the lower amplitude desired channel.

INTERMODULATION

Intermodulation products (IMPs) are created when signals are present or operated in the non-linear region of an active device such as an amplifier or a mixer. The third-order products are typically of greatest concern, since they are lowest order that can fall within the passband of a sub-octave bandwidth receiver. The presence of intermodulation products can be created by two primary mechanisms in a co-located transmitter and receiver environment. The first is receiver IMPs created within the front-end circuitry of the receiver. Consider the following VHF system,



which represents the front-end circuitry of the basic VHF communications receiver in the field today. The optional external bandpass filter is a single pole cavity with an nominal 3 dB bandwidth of 0.100 MHz. This is installed in front of receivers at locations with known interference problems. A frequency response plot is provided in figure B-1. The internal preselector filter is a tunable 2-pole helical resonator bandpass filter with a 3 dB bandwidth of 1.1 MHz. The frequency response plot is provided in figure B-2. The input power of an interfering signal which will generate intermodulation products at levels above

the receiver noise floor can be predicted by the RF bandwidth of the receiver, the noise figure, the gain,and the intercept point. Since the front-end device characteristics dominate the overall receiver performance specifically noise figure and intercept point, many interference mechanisms can be highlighted with the simple model above.

Consider the following conditions:

RF bandwidth — 1 MHz
front-end gain (G) — 10 dB
third-order intercept point (IP_3) — 14 dBm
noise figure (NF) — 15 dB

the input noise floor within the RF bandwidth can be calculated by the following relationship,

$$N_{RF} = NF - 114\ \text{dBm/1 MHz} + 10 \cdot \log(BW_{RF})$$

which for the above conditions is,

$$N_{RF} = -99\ \text{dBm}$$

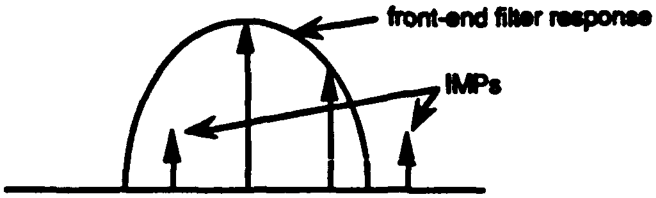
The level at which two in-band signals will create a third-order intermodulation product equal to N_{RF} can be predicted by,

$$P = \frac{1}{3}(N_{RF} - 2 \cdot G + 2 \cdot IP_3)$$

where this equals -30 dBm for the above conditions. The third-order products will be at frequencies,

$$2f_2 \pm f_1 \text{ and } 2f_1 \pm f_2$$

The levels of the intermodulation products will depend on the input levels of each primary input frequency to the receiver front-end. One common scenario for a receiver in a multiple signal environment is the desired signal centered in the passband and the undesired signal is only partially attenuated by front-end filter or preselector. This is graphically illustrated below.



The levels of the intermodulation products can be approximated by the following relationships for signal frequencies f_1 and f_2 at input power levels P_1 and P_2 respectively.

$$IM_3 = 2 \cdot P_1 + P_2 - 2 \cdot IP_3 + 3 \cdot G \quad \text{for } 2 \cdot f_1 \pm f_2$$
$$= 2 \cdot P_2 + P_1 - 2 \cdot IP_3 + 3 \cdot G \quad \text{for } 2 \cdot f_2 \pm f_1$$

Clearly the receiver intercept point and total gain determines the level of the intermodulation products. Certain components within a receiver front-end can create a varying intercept point with input level. The presence of very strong signals in components such as varactor or PIN diodes can change their characteristics. For most fixed or manually tuned receivers, this is not a problem. Frequency hopped or switched filters rely on such circuit elements and would have additional concerns at high signal levels.

Transmitter Back Intermodulation Products

One source of unwanted products in a multiple transmitter environment is BIPs. The signal from one co-located transmitter can be received by the antenna of an adjacent transmitter and pass into the power amplification circuits. The two signals in the amplifier intermodulate and transmit both fundamental and harmonically related products. This effect is a function of the reverse isolation, the class of the amplifier, and the signal power levels. The level of the IMPs follows the basic relationships above, where the fundamental signal levels are determined by the antenna-to-antenna isolation, the reverse isolation of the amplifier, and the forward gain of the amplifier, while the IMP levels are determined by the linearity of the amplifier.

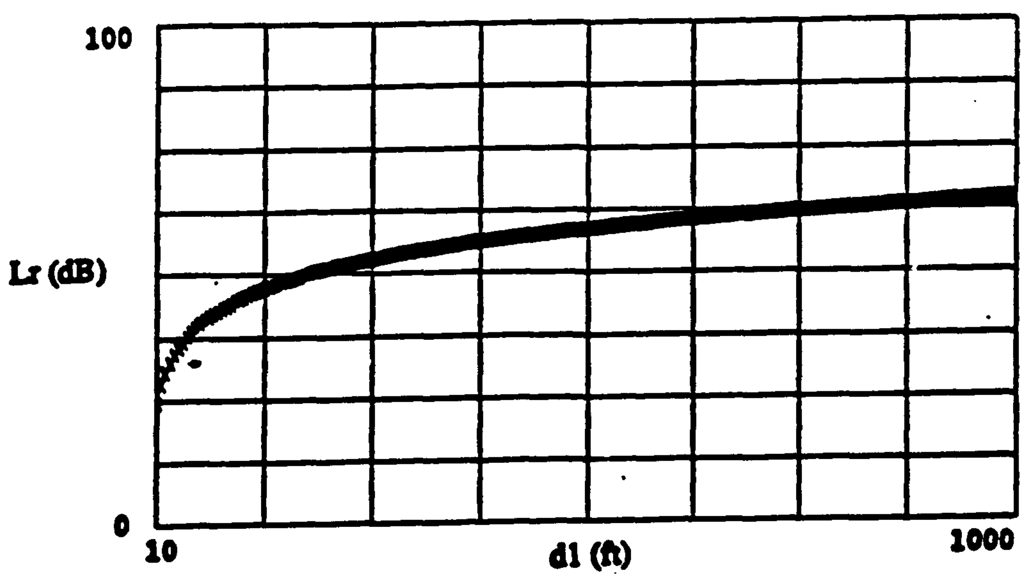
Densensitization/Cross Modulation

Desensitization is a reduction in the receiver gain to the desired signal as a result of an interfering emission producing automatic-gain control (AGC) action or causing one or more active amplification stages to operate above saturation. Cross-modulation is the transfer of

modulation from an undesired emission onto the desired signal due to non-linearities in the receiver. The influence of strong undesired signals on the receiver's performance for each of these interference mechanisms is related to the linearity of the front-end components and the amount of front-end selectivity. However a multi-signal cosite environment is rarely used to specify the performance of these two variables.

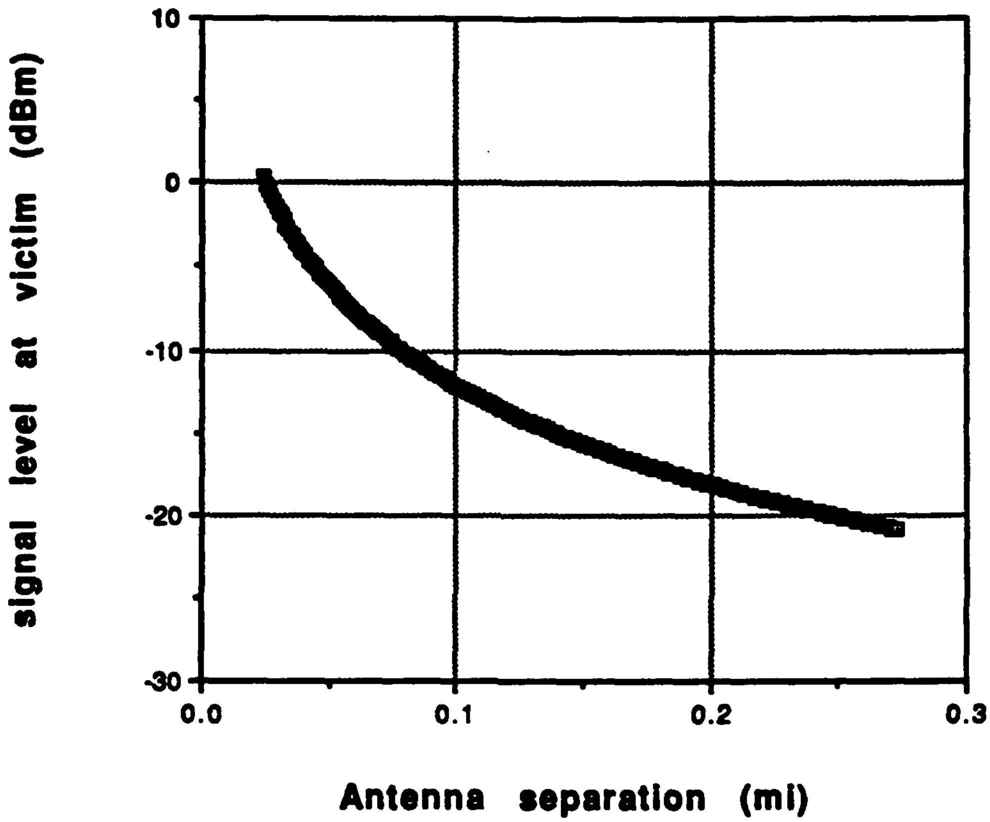
Site Configurations

The amount of isolation between VHF transmitters and receivers is a function of the site configuration. ATC VHF antennas can be placed on several colinear masts on top of the control tower or placed separately on the ground based masts. Separations can range from 10 to 100s of feet. The nominal output power is either 10 W or 50 W with omni-directional antenna elements. VHF propagation falls in the reflection zone, where a received signal is a composite of both direct and reflected waves. The reflected component becomes significant at antenna heights greater than 10λ. A conservative transmission model for interfering signals can assume free-space propagation, and nominal losses and gains for the cables and antennas. Free-space loss is plotted below for a mid-band frequency of 127.5 MHz at separations ranging from 10 to 1000 feet.



Assigning 2.5 dB gain for each antenna and 5.0 dB loss for cables over the entire link, the above free-space attenuation plot would be valid for signal loss between a given

transmitter and a separate receiver. Consider a 50 W (47 dBm) transmitter, the received signal strength as a function of antenna separation based on free-space propagation is displayed below.



The level of an undesired VHF frequency in the presence of a desired frequency, that would yield an intermodulation product at the receiver noise floor, was previously stated as being approximately -30 dBm for the basic receiver front-end model. The undesired frequency would need to be offset from the desired frequency by at least 1.75 MHz to avoid potential interference from intermodulation products and desense, given the current internal preselector characteristics.

Adjacent Channel Considerations

The current ground VHF ATC receiver currently suppresses adjacent channel signals, 25 kHz away, approximately 60 dB at the final IF. Tests have shown that a 0.5 to 0.6 nmi separation is needed given a 10 W interfering transmitter. To maintain this same level of frequency channel protection for equal receiver performance, the same level of suppression must be provided in a closer channelized receiver. Either the combined receiver filter suppression must offer in excess of 60 dB ± 12.5 kHz or the close-in spurious and noise must be improved for all co-located transmitters. The latter is not practical since the receiver suppression is achievable.

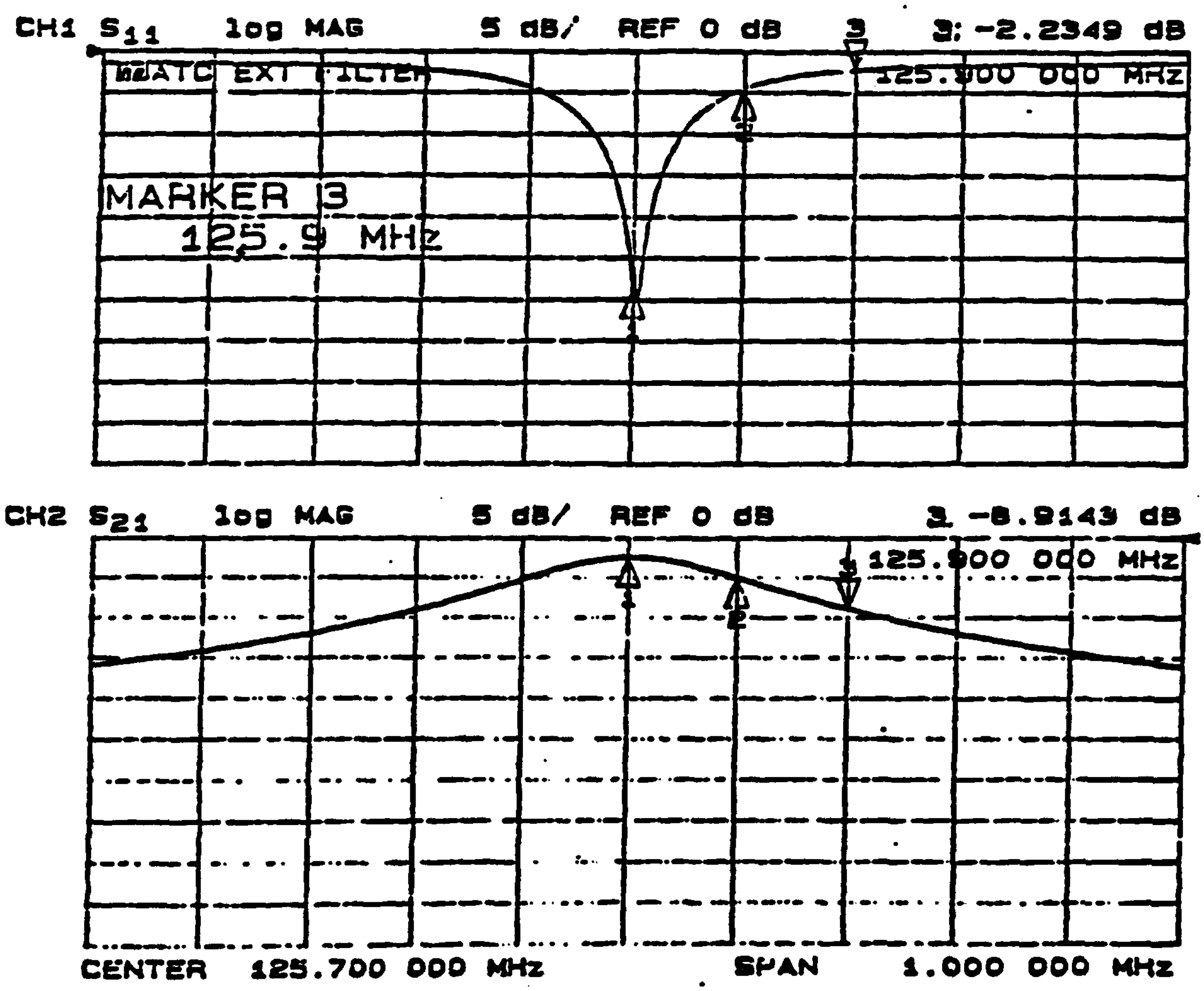
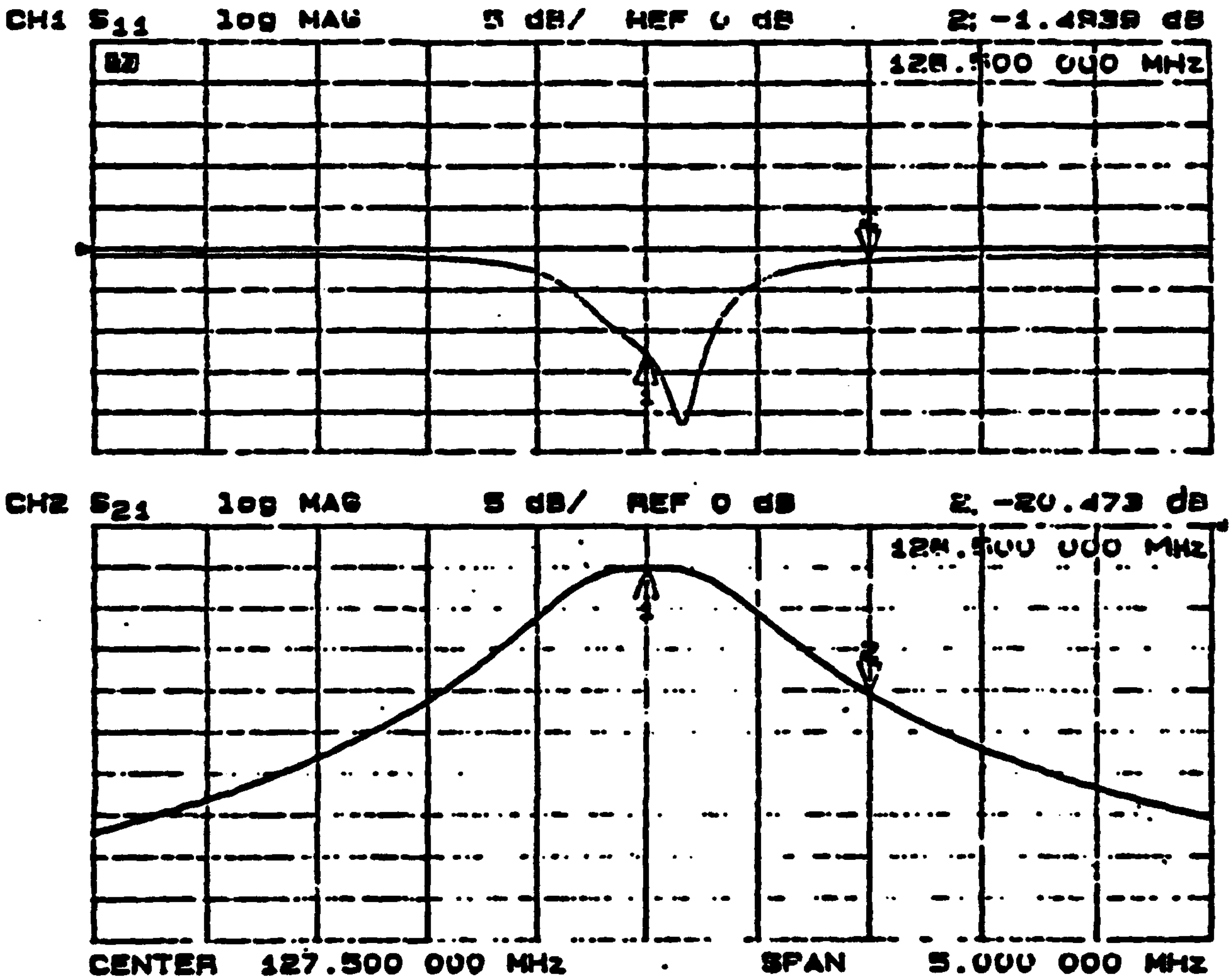


Figure B-1



CH1 S11 log MAG 5 dB/ REF 0 dB 2: -1.4939 dB
128.500 000 MHz
CH2 S21 log MAG 5 dB/ REF 0 dB 2: -20.473 dB
128.500 000 MHz
CENTER 127.500 000 MHz SPAN 5.000 000 MHz

Figure B-2

GLOSSARY

ACARS	Aircraft Communications Addressing and Reporting System
ACI	Adjacent Channel Interference
ACK	Acknowledgment
ADS	Automatic Dependent Surveillance
AEEC	Airlines Electronic Engineering Committee
A/G	Air/Ground
ALOHA	A simple random access technique originated by the University of Hawaii
AM	Amplitude Modulation
AMCP	Aeronautical Mobile Communications Panel
AME	Amplitude Modulation Equivalent
AOC	Aeronautical Operational Control
ARINC	Aeronautical Radio, Incorporated
A-QPSK	Aviation QPSK
ARQ	Automatic Repeat Request
ATC	Air Traffic Control
ATIS	Automatic Terminal Information Service
ATM	Air Traffic Management
ATN	Aeronautical Telecommunications Network
ATS	Air Traffic Services
AVPAC	Aviation VHF Packet Communications
BCH	Bose-Chaudhuri-Hocquenghem
BER	Bit Error Rate
BFSK	Binary Frequency Shift Keying
BPSK	Binary Phase Shift Keying
BUEC	Backup Emergency Communications
CA	Commercial Aviation
CAASD	Center for Advanced Aviation System Development
CD	Collision Detection
CDM	Code Division Multiplexing
CDMA	Code Division Multiple Access
CELP	Codebook Excited Linear Prediction
COM	Communications
CRA	Collision Resolution Algorithm
CRI	Collision Resolution Interval
CSMA	Carrier Sense Multiple Access
CTAG	Cellular Trunked Air/Ground
CVSD	Continuously Variable Slope Delta (Modulation)
DSBSC	Double Sideband Suppressed Carrier
DSBTC	Double Sideband Transmitted Carrier
FAA	Federal Aviation Administration
FDM	Frequency Division Multiplexing
FDMA	Frequency Division Multiple Access

FEC	Forward Error Correction
FIFO	First-In First-Out
FM	Frequency Modulation
FSK	Frequency Shift Keying
4LFM	Four Level FM
GA	General Aviation
GMSK	Gaussian Minimum Shift Keying
ICAO	International Civil Aviation Authority
ICC	International Conference on Communications
IEEE	Institute of Electrical and Electronics Engineers
IF	Intermediate Frequency
IMBE	Imrproved Multi-Band Excitation
IMP	Intermodulation Product
INT	Integer
ISO	International Standards Organization
LAK	Lack (of) Acknowledgment
LBT	Listen Before Talk
LO	Local Oscillator
LPC	Linear Predictive Coding
MHz	MegaHertz
MIL	Military
MSK	*Minimum Shift Keying*
MSR	MITRE Sponsored Research
N/A	Not Applicable
NAK	Negative Acknowledgment
NAS	National Airspace System
NBFM	Narrowband Frequency Modulation
OSI	Open Systems Interconnection
OQAM	Offset Quadrature Amplitude Modulation
PDC	Pre-Departure Clearance
PSK	Phase Shift Keying
PTT	Push To Talk
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RA	Random Access
RF	Radio Frequency
RS	Reed-Solomon
RTCA	Radio Technical Commission for Aeronautics (only initials, not former name, used now)

SC	Special Committee
SSB	Single Sideband
STD	Standard
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
UK CAA	United Kingdom Civil Aviation Authority
VDR	VHF Data Radio
VHF	Very High Frequency
VSELP	Vector Sum Code Excited Linear Predictive (Coding)
VSB	Vestigial Sideband
WG	Working Group
WP	Working Paper